JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 34, 103-121 (2018)
DOI: 10.6688/JISE.2018.34.1.7

A Fast and Automatic Kernel-based Classification Scheme: GDA+SVM or KNWFE+SVM*

CHENG-HSUAN LI[1,*], PEI-JYUN HSIEN[1] AND LI-HUI LIN[2,3]
[1]*Graduate Institute of Educational Information and Measurement*
National Taichung University of Education
Taichung, 40306 Taiwan
[2]*College of Mathematics and Computer Science*
Wuyi University
Wuyishan, 354300 P.R. China
[3]*The Key Laboratory of Cognitive Computing and Intelligent Information Processing*
Fujian Education Institutions
Wuyishan, 354300 P.R. China
E-mail: {chenghsuanli; pp110349}@gmail.com; 6390984@qq.com

For high-dimensional data classification such as hyperspectral image classification, feature extraction is a crucial pre-process for avoiding the Hughes phenomena. Some feature extraction methods such as linear discriminant analysis (LDA), nonparametric weighted feature extraction (NWFE), and their kernel versions, generalized discriminant analysis (GDA) and kernel nonparametric weighted feature extraction method (KNWFE) have been shown that they can improve the classification performance. However, there are two challenges which influence the classification performance by applying GDA or KNWFE. First one is the solution of the generalized eigenvalue problem formed by "implicit" within- and between-class scatter matrices. The other one is the appropriate selection of the kernel parameter(s). Therefore, researchers rarely implement them for dealing with high-dimensional data classification. Recently, an automatic kernel parameter selection method (APS) was proposed to predetermine the appropriate RBF kernel for support vector machine (SVM) instead of the transitional cross-validation method. In this study, a theoretical procedure to solve the implicit generalized eigenvalue problem was proposed. Moreover, APS was applied to find the suitable RBF kernel parameter of GDA and KNWFE. Combing with kernel-based classifier, SVM, a fast and automatically kernel-based classification scheme, GDA+SVM or KNWFE+SVM, was also brought out. From the experiment results on real data sets, the classification performance of GDA+ SVM or KNWFE+SVM outperforms SVM with whole features, especially in the small sample size problem. Most importantly, the readership can extend any feature extraction methods based on within- and between-class scatter matrices. In addition, the researcher can implement them directly without tuning the kernel parameter.

Keywords: kernel method, feature extraction, variable selection, GDA, KNWFE

1. INTRODUCTION

Feature extraction is an important step for high dimensional classification problem due to the small sample size problem [3-7, 21]. Moreover, it can eliminate irrelevant features and construct more relevant features for classification [20]. For example, in hyperspectral image classification, it suggests that transforming data into a lower dimensional space before classification [10]. Traditionally, linear discriminant analysis (LDA)

Received January 5, 2016; revised June 16, 2016 & June 20, 2017; accepted July 26, 2017.
Communicated by Shou-De Lin.

[1] and nonparametric weighted feature extraction method (NWFE) [7] were used for dimension reduction. Lots of researchers have shown that doing LDA or NWFE can improve the classification performance in hyperspectral image classification problem [1, 3, 6-9]. However, the transformations of LDA and NWFE are linear because they use a transformation matrix to reduce the dimensionality.

The kernel trick is a typical and an easy way to extend the linear model to nonlinear model [4, 6, 8, 9]. In nonlinear support vector machine (SVM) [9, 11], kernel functions such as the polynomial kernel, the radial basis function (RBF) kernel, or the sigmoid kernel are instead of the inner product terms in the objective function and constraints of linear SVM. Hence, the linear hyperplane in the kernel-induced feature space [9] is with respect to a nonlinear boundary in the original space. The kernel trick can be applied to extend the linear feature extraction methods into nonlinear feature extraction methods [6, 8, 9, 11]. The kernel principal component analysis (KPCA) [11], generalized discriminant analysis (GDA) [8], and kernel nonparametric weighted feature extraction method (KNWFE) [6, 9] are the kernel-based feature extraction methods of principal component analysis (PCA) [1], LDA [1], and NWFE [7], respectively. Moreover, Kuo *et al.* showed that KNWFE outperforms other feature extraction methods [6].

However, the solution of the generalized eigenvectors of GDA [8] and KNWFE [9] are not accurate for the real applications. If the kernel matrix is nearly singular, some generalized eigenvectors with respect to the eigenvalues whose values are close to 0 should be eliminate. Hence, the multiplication of the transpose of the matrix formed by the generalized eigenvectors with respect to nonzero (far from 0) eigenvalues ant itself is not equal to 0. There is an example shown this circumstance in this study. Hence, how to find more accurate generalized eigenvector becomes a crucial problem by applying the kernel-based feature extraction method based on implicit within- and between-class scatter matrices. Here, we proposed a theoretical process for solving the generalized eigenvalue problem. The proposed method does not involve the eigenvalue decomposition of the kernel matrix and the assumption of the dual form. That is, the generalized eigenvectors can be computed directly.

The other big issue by applying kernel-based algorithms is on kernel parameter selection [9, 12, 13]. The poor parameter decreases the classification performance. An automatic kernel parameter selection method (APS) was proposed to determine automatically the RBF kernel parameter instead of the traditional cross-validation method before applying SVM or GDA with RBF kernel [12-14]. Moreover, the classification performance of GDA+SVM, SVM with GDA feature subsets, still outperforms that of SVM with whole features. The experimental results are similar to those shown in the paper. In addition, APS speeds up the training time of the kernel-based classification scheme, kernel-based feature extraction method (GDA) + kernel-based classifier (SVM).

Therefore, in this study, APS [9, 12, 13] is applied to predetermine the appropriate RBF kernel parameter of GDA and KNWFE. Then based on the parameter, the proposed theoretical process is used to compute the generalized eigenvectors and the projected vectors of GDA and KNWFE directly without involving the eigenvalue decomposition of the kernel matrix. Moreover, the SVM with APS is applied for classification based on projected samples with lower dimensionality. Therefore, a fast and automatic kernel-based classification scheme, GDA+SVM or KNWFE+SVM with APS, was proposed. The RBF kernel parameters of GDA, KNWFE, and SVM are automatically determined

by APS in few seconds. The related work, APS, GDA, and KNWFE are introduced in section 2. The section 3 describes the theoretical process to find the generalized eigenvectors of GDA and KNWFE. Moreover, the flowchart of GDA+SVM and KNWFE+SVM with APS is also described. The details of experimental results on real data sets and the conclusion are illustrated in sections 4 and 5, respectively.

2. RELATED WORK

In this study, the APS is used to pre-determine the suitable RBF kernel parameter of GDA and KNWFE. Hence the APS, GDA, and KNWFE are introduced in this section.

2.1 Automatic Kernel Parameter Selection Method

Suppose the training set is

$$\{x_j^{(i)}\}_{j=1,..,N_i} \subset R^d$$

where j means the jth samples in class i, i = 1, 2, …, L, and N_i indicates the number of training samples in the ith class. In automatic kernel parameter selection (APS) method, the within- and between-class separability measures were used to compute and compare class separabilities with respect to different RBF kernel parameters. The within- and between-class separability measures are defined by

$$w(\sigma) = \frac{1}{\sum_{i=1}^{L} N_i^2} \sum_{i=1}^{L} \sum_{\ell=1}^{N_i} \sum_{k=1}^{N_i} \kappa(x_\ell^{(i)}, x_k^{(i)}, \sigma)$$

and

$$b(\sigma) = \frac{1}{\sum_{i=1}^{L} \sum_{\substack{j=1 \\ j \neq i}}^{L} N_i N_j} \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j \neq i}}^{L} \sum_{\ell=1}^{N_i} \sum_{j=1}^{N_j} \kappa(x_\ell^{(i)}, x_k^{(j)}, \sigma)$$

respectively, where

$$\kappa(x, x', \sigma) = \exp\left(-\frac{\| x - x' \|^2}{2\sigma^2} \right), \; x, x' \in R^d, \; \sigma \in R - \{0\}$$

is the RBF kernel function. Moreover, the class separability can be computed by

$$J(\sigma) = 1 - w(\sigma) + b(\sigma).$$

Then, the appropriate RBF kernel parameter is a value σ such that $J(\sigma)$ is maximal. That is, the within-class separability is close to 1 and the between-class separability is close to 0, simultaneously. Hence, in the corresponding feature space, the samples in the same class are close enough, and, in the same time, the samples from different classes are far enough away. APS has been applied to SVM [12, 13] and GDA [14]. The experimental results showed that APS can find a suitable parameter in few seconds (faster than the cross-validation method) and improve the class performance [14].

2.2 Generalized Discriminant Analysis

Generalized discriminant analysis (GDA) was proposed by Baudat and Anouar [19]. The between-class and with-class scatter matrices, S_b^{GDA} and S_w^{GDA} in the feature space are defined as

$$S_b^{GDA} = \sum_{i=1}^{L} \frac{N_i}{N} (m_i - m_0)(m_i - m_0)^T$$

and

$$S_w^{GDA} = \frac{1}{N} \sum_{i=1}^{L} \sum_{j=1}^{N_i} (\phi(x_j^{(i)}) - m_i)(\phi(x_j^{(i)}) - m_i)^T$$

where

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(x_j^{(i)}),\ m_0 = \frac{1}{N} \sum_{i=1}^{L} \sum_{j=1}^{N_i} \phi(x_j^{(i)}), \text{ and } N = N_1 + \ldots + N_L.$$

Based on some matrix computation,

$$S_b^{GDA} = X^T B_{GDA} X \text{ and } S_w^{GDA} = X^T B_{GDA} X,$$

where $X^T = [X_1^T, \ldots, X_L^T]$, $X_i^T = [\phi(x_1^{(i)}), \ldots, \phi(x_{N_i}^{(i)})]$, $i = 1, 2, \ldots, L$,

$$W = \frac{1}{N} \left\{ I_N - \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_L \end{bmatrix} \right\},\ B = \frac{1}{N} \left\{ \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_L \end{bmatrix} - \begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix} \right\}, \text{ and}$$

$$W_i = \begin{bmatrix} \frac{1}{N_i} & \cdots & \frac{1}{N_i} \\ \vdots & \ddots & \vdots \\ \frac{1}{N_i} & \cdots & \frac{1}{N_i} \end{bmatrix}.$$

Note that the feature mapping ϕ is implicit, and so are the scatter matrices. Hence, the transformation matrix cannot be found by solving the following generalized eigenvalue problem

$$S_b^{GDA} v = \lambda S_w^{GDA} v.$$

2.3 Kernel Nonparametric Weighted Feature Extraction

Kernel nonparametric weighted feature extraction (KNWFE) is a kernel version of nonparametric weighted feature extraction (NWFE) [6, 9]. The within-class scatter matrix S_w^{KNWE} and the between-class scatter matrix S_b^{KNWE} of KNWFE in the feature space are

defined as

$$S_w^{KNWE} = \sum_{i=1}^{L} P_i \sum_{\ell=1}^{N_i} \frac{\lambda_\ell^{(i,i)}}{N_i} (\phi(x_\ell^{(i)}) - M_i(\phi(x_\ell^{(i)})))(\phi(x_\ell^{(i)}) - M_i(\phi(x_\ell^{(i)})))^T$$

and

$$S_b^{KNWE} = \sum_{i=1}^{L} P_i \sum_{\substack{j=1 \\ j \neq i}}^{L} \sum_{\ell=1}^{N_i} \frac{\lambda_\ell^{(i,i)}}{N_i} (\phi(x_\ell^{(i)}) - M_j(\phi(x_\ell^{(i)})))(\phi(x_\ell^{(i)}) - M_j(\phi(x_\ell^{(i)})))^T$$

where the scatter matrix weight $\lambda_\ell^{(i,j)}$ is defined by

$$\lambda_\ell^{(i,j)} = \frac{\| \phi(x_\ell^{(i)}) - M_j(\phi(x_\ell^{(i)}) \|^{-1}}{\sum_{t=1}^{N_i} \| \phi(x_t^{(i)}) - M_j(\phi(x_t^{(i)}) \|^{-1}},$$

$M_j(\phi(x_\ell^{(i)})) = \sum_{k=1}^{N_j} w_{\ell k}^{(i,j)} \phi(x_k^{(j)})$ denotes the weighted mean with respect to $\phi(x_\ell^{(i)})$ in class j and

$$w_{\ell k}^{(i,j)} = \frac{\| \phi(x_\ell^{(i)}) - \phi(x_k^{(j)}) \|^{-1}}{\sum_{t=1}^{N_j} \| \phi(x_t^{(i)}) - \phi(x_t^{(j)}) \|^{-1}}.$$

Although KNWFE outperforms GDA, NWFE, LDA, DBFE (decision-boundary feature extraction), and *etc*. [6], it is still hard applied for readership due to the kernel parameter selection.

Similar to GDA, the scatter matrices S_b^{KNWFE} and S_w^{KNWFE} can be decomposed into $S_b^{KNWFE} = X^T B_{KNWFE} X$ and $S_w^{KNWFE} = X^T W_{KNWFE} X$, respectively, where the formulations of B_{KNWFE} and W_{KNWFE} can be found in our previous paper. So, the corresponding transformation matrix is formed by the generalized eigenvectors of

$$S_b^{KNWFE} v = \lambda S_w^{KNWFE} v.$$

2.4 Eigenvalue Resolution with Regularization

Both GDA and NWFE, the between- and within-class scatter matrices are the forms

$S_b = X^T B X$ and $S_w = X^T W X$,

respectively. In GDA, $B = B_{GDA}$ and $W = W_{GDA}$. Moreover, $B = B_{KNWFE}$ and $W = W_{KNWFE}$ in KNWFE. The goal is to find the transformation matrix $A = [v_1, v_2, \ldots, v_p] \in R^{d \times p}$ such that

$S_b v_i = \lambda_i S_w$, $v_i = 1, 2, \ldots, p$ with $\|v_i\| = 1$ and $\|v_i\| = 1$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

In general, it is to find the eigen-pairs (λ_i, v_i) of $S_w^{-1} S_b$ in the feature space. However, the scatter matrices are implicit. Hence, it is impossible to find the eigen-pairs from scat-

ter matrices directly. Hence, the dual form assumption is used to find the eigen-pairs [8, 9].

Suppose the dual form of the generalized eigenvector v_i is

$$v_i = X^T \bar{v}_i,\ \bar{v}_i \in R^{N_i \times 1}.$$

Then

$$KBK\bar{v}_i = \lambda_i KBK\bar{v}_i,\ i = 1, 2, \ldots, p \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p, \tag{1}$$

where $K = XX^T$ is an $N \times N$ kernel matrix. If $N > d$, the kernel matrix is singular, and so are KBK and KWK.

Hence, the eigen-decomposition of the kernel matrix K is employed to overcome this problem [8, 9]. Assume $K = P\Gamma P^T$, where Γ is a square matrix whose diagonal entries are nonzero eigenvalues of K and other entries are zeros, and the columns of P are the eigenvectors of K with respect to the nonzero eigenvalues. Then we can solve the eigen-pairs of

$$P^T BP\tilde{v}_i = \lambda_i P^T WP\tilde{v}_i,\ i = 1, 2, \ldots, p \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p.$$

Similarly, $(\lambda_i, \tilde{v}_i)$ is the eigen-pairs of $(P^T WP)^{-1} P^T BP$. Nevertheless, the estimation of $(P^T WP)^{-1}$ may be nearly singular. Hence, a regularization,

$$0.5P^T WP + 0.5\text{diag}(P^T WP),$$

is taken for decreasing the influence of the singularity [9].

After finding $\tilde{v}_i$, the

$$\bar{v}_i = P\Gamma^{-1}\tilde{v}_i. \tag{2}$$

In addition, the column vectors of the transformation matrix A is

$$v_i = X^T \frac{\bar{v}_i}{\sqrt{\bar{v}_i^T K \bar{v}_i}}.$$

Finally, the projected components of an unknown sample $\phi(z)$ in the feature space can be calculating by

$$\phi(z)^T v_i = \sum_{j=1}^{L} \sum_{\ell=1}^{N_j} \frac{\bar{v}_i}{\sqrt{\bar{v}_i^T K \bar{v}_i}} \kappa(x_\ell^{(j)}, z), \text{ for } i = 1, 2, \ldots, p.$$

3. PROPOSED METHOD

There are two important issues for applying GDA and KNWFE before using SVM. The first one is how to find the more accurate generalized eigenvectors, and the other one is the appropriate kernel parameters. The theoretical process for solving the generalized eigenvalue problem based on implicit scatter matrices was proposed in this section. Moreover, the APS was used to determine the suitable RBF kernel parameter of GDA

and KNWFE in few seconds.

3.1 Theoretical Process for Solving the Generalized Eigenvalue Problem

The procedure illustrated in Section 2.4 has some drawbacks on estimations. First one, Eq. (1) takes two K. It increases the singularity for solving the generalized eigenvalue problem. Second, the eigen-decomposition is applied to decrease the influence of the singularity. However, taking one more eigen-decomposition may affect estimations of the eigenvectors. Then, how may nonzero eigenvlaues of K should be considered for constructing Γ? It is also a problem in Eq. (2), if $\Gamma$is nearly singular, it may influence the solution, too. The most important one is taking P in Eq. (2). If all eigenvlaues of $K$are nonzero, then we have

$$P^T P = PP^T = I,$$

where I is an identity matrix. However, if only first few eigenvalues of K are considered, then $P^T P$ is still an identity matrix, but PP^T is not an identity matrix. For example,

$$K = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} = P\Gamma P^T,$$

where

$$P = \begin{bmatrix} 0.1826 & 0.8944 & 0.4082 \\ 0.3651 & -0.4472 & 0.8165 \\ -0.9129 & 0 & 0.4082 \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 12 \end{bmatrix}.$$

Note that $P^T P \approx I$ and $PP^T \approx I$. However, if we consider the first two nonzero eigenvalues and the corresponding eigenvectors, that is,

$$P = \begin{bmatrix} 0.8944 & 0.4082 \\ -0.4472 & 0.8165 \\ 0 & 0.4082 \end{bmatrix},$$

then

$$P^T P \approx \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ but } PP^T \approx \begin{bmatrix} 0.9667 & -0.0667 & 0.1667 \\ -0.0667 & 0.8667 & 0.3333 \\ 0.1667 & 0.3333 & 0.1667 \end{bmatrix}$$

is very different from the 3×3 identity matrix. Therefore, $\tilde{v}_i = \Gamma P^T \bar{v}$ cannot imply

$$\bar{v} = P\Gamma^{-1}\tilde{v}_i,$$

directly, if only few eigenvectors are considered.

In this study, we introduced an alternative procedure, a theoretical process, to find the more accurate generalized eigenvector for the kernel-based feature extraction method based on the implicit within- and between-class scatter matrices. Recall that, both GDA and NWFE, the between- and within-class scatter matrices are the forms

$$S_b = X^T B X \text{ and } S_w = X^T W X,$$

respectively. And the goal is to find the transformation matrix $A = [v_1, v_2, \ldots, v_p] \in R^{d \times p}$ such that

$$X^T B X v_i = \lambda_i X^T W X v_i,\ i = 1, 2, \ldots, p \text{ with } \|v_i\| = 1 \text{ and } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p.$$

First, find the generalized eigenvectors of

$$BKu_i = \lambda_i WKu_i,\ i = 1, 2, \ldots, p \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p. \tag{3}$$

Then multiplying both sides of Eq. (3), we have

$$X^T B X X^T u_i = \lambda_i X^T W X X^T u_i,\ i = 1, 2, \ldots, p \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p.$$

Hence, after finding the generalized eigenvectors, u_i, the normalized generalized eigenvectors,

$$v_i = X^T \frac{u_i}{\sqrt{u_i^T K u_i}},$$

can be calculated directly without other eigenvalue decomposition or taking inverse of some near singular matrix. Therefore, the projected components of an unknown sample $\phi(z)$ in the feature space can be calculating by the following two steps.

Step 1: Find the generalized eigenvectors, u_i, of

$$BKu_i = \lambda_i WKu_i,\ i = 1, 2, \ldots, p \text{ with } \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p.$$

Step 2: The ith projected component of an unknown sample $\phi(z)$ is

$$\phi(z)^T v_i = \sum_{j=1}^{L} \sum_{\ell=1}^{N_j} \frac{u_i}{\sqrt{u_i^T K u_i}} \kappa(x_\ell^{(j)}, z) \cdot$$

3.2 GDA and KNWFE with APS

Similar to SVM with APS [13], the APS is applied to predetermine the RBF kernel parameter of GDA [14] and KNWFE. Then the reduced data is classified by SVM with APS. Fig. 1 illustrates the flowchart of GDA+SVM or KNWFE+SVM with APS. First, the appropriate RBF kernel parameter of GDA or KNWFE is pre-determined by APS. Then, the transformation $A \in R^{d \times p}$ with respect to GDA or KNWFE by solving the generalized eigenvalue problem

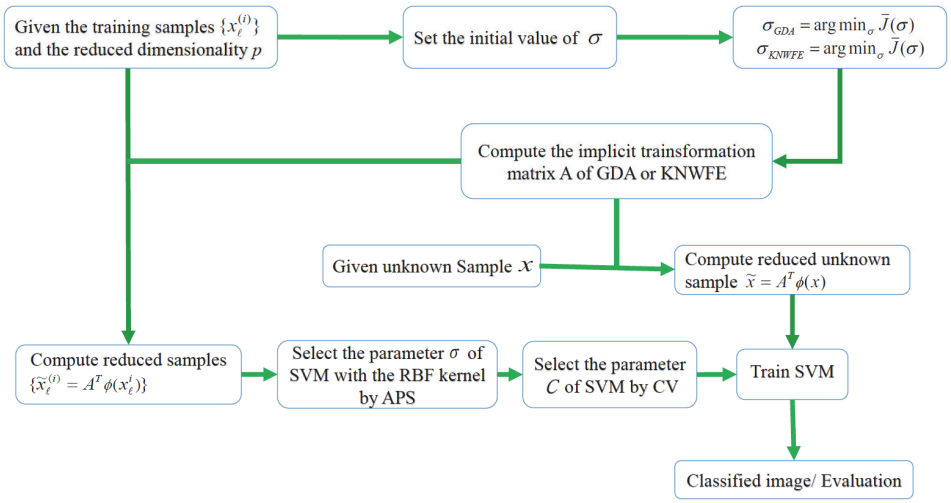

Fig. 1. Proposed kernel-based classification scheme, GDA+SVM or KNWFE+SVM with APS.

$$S_b v = \lambda S_w v$$

based on proposed procedure in Section 3.1. Moreover, compute the reduced training samples $\{A^T x_j^{(i)}\}_{j=1,..,N_i} \subset R^p$.

After obtaining the reduced training samples, train the SVM with the reduced training samples based on APS. The penalty parameter is determined by cross-validation method. (According to our previous paper [13], if the RBF kernel parameter is appropriate, the penalty parameter can be determined by try and error in few times.) Finally, use the trained SVM with APS to classify the reduced unknown sample $A^T x$ where $x \in R^d$ is an unknown sample in the original space. Note that the matrix product computations or solving generalized eigenvalue problem should be calculated according to the kernel trick.

The steps of the proposed approaches are listed as follows:

Step 1: Given the training samples $\{x_j^{(i)}\}_{j=1,..,N_i} \subset R^d$ and the reduced dimensionality p.
Step 2: Use the training samples $\{x_j^{(i)}\}_{j=1,..,N_i} \subset R^d$ to find the best parameter σ_{GDA} (σ_{KNWFE}) to minimize the following equation with the given training samples:

$$\min_\sigma J(\sigma) = 1 - \omega(\sigma) + b(\sigma) \cdot$$

Step 3: Compute the implicit transformation matrix A of GDA (KNWFE).
Step 4: Find the reduced training samples $\{\tilde{x}_j^{(i)}\}_{j=1,..,N_i} \subset R^p$ by applying GDA (KNWFE) based on the RBF kernel function with the optimal parameter σ_{GDA} (σ_{KNWFE}).
Step 5: Use the reduced training samples $\{\tilde{x}_j^{(i)}\}_{j=1,..,N_i} \subset R^p$ to find the best parameter σ to minimize the following equation with the reduced training samples:

$$\min_\sigma J(\sigma) = 1 - \omega(\sigma) + b(\sigma) \cdot$$

Step 6: Use k-fold cross-validation (CV) to find appropriate parameter C, the penalty parameter of SVM, of the soft-margin SVM by applying the RBF kernel function κ with σ.
Step 7: Compute the transformed unknown sample $\tilde{x} \in R^p$ of a given unknown sample $x \in R^d$ by applying GDA (KNWFE) based on the RBF kernel function with the optimal parameter σ.
Step 8: Use the trained soft-margin SVM with σ and C to classify the transformed unknown sample $\tilde{x}$.

4. EXPERIMENTAL RESULTS

This study compared the classification performance of SVM with original features [11], and feature subset of GDA [8] and KNWFE [6] on some real data sets. Moreover, the RBF kernel parameter of GDA, KNWFE, and SVM were determined by APS, automatically.

4.1 Real Data Sets

This study used two real hyperspectral data sets and two UCI data sets to evaluate the classification performance of the fast kernel classification scheme, GDA+SVM and KNWFE+SVM based on APS. For hyperspectral image data sets, the Indian Pine Site (IPS) and the Pavia University (PAVIA) data set were considered. The IPS is a mixed forest/agricultural site in Indian [10] and was gathered by a sensor known as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). This data set was obtained from an aircraft operated by the NASA/Jet Propulsion Laboratory and flown at an altitude of 65 000 ft. The images measured 145×145 pixels, with 220 spectral bands and a 20m spatial resolution. Fig. 2 shows the IR image and ground truth of IPS. Sixteen land-cover classes were available in the original ground-truth image: Alfalfa (Class 1), Corn-no till (Class 2), Corn-min till (Class 3), Corn (Class 4), Grass-pasture (Class 5), Grass-trees (Class 6), Grass-pasture-mowed (Class 7), Hay-windrowed (Class 8), Oats (Class 9), Soybean-notill (Class 10), Soybean-mintill (Class 11), Soybean-clean (Class 12), Wheat (Class 13), Woods (Class 14), Buildings-Grass-Trees-Drives (Class 15), and Stone-Steel-Towers (Class 16). Table 1 lists the number of pixels of each class.

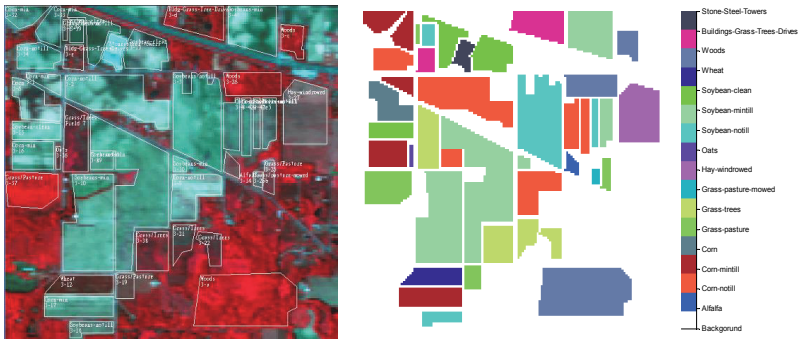

Fig. 2. A portion of the Indian Pine Site image with a size of 145×145 pixels and the ground truth of the Indian Pine Site data set.

Table 1. Sixteen categories and corresponding number of pixels in the Indian Pine Site image. #(pixels) indicates the number of pixels per class.

No.	Category	#(pixels)	No.	Category	#(pixels)
1	Alfalfa	46	9	Oats	20
2	Corn-no till	1428	10	Soybeans-notill	972
3	Corn-min till	830	11	Soybeans-mintill	2455
4	Corn	237	12	Soybeans-clean	593
5	Grass-pasture	483	13	Wheat	205
6	Grass-trees	730	14	Woods	1265
7	Grass-pasture-mowed	28	15	Buildings-Grass-Trees-Drives	386
8	Hay-windrowed	478	16	Stone-Steel-Towers	93

For investigating the influence of training sample size to the dimensionality, some classes with few samples were ignored in this study. Only twelve classes, Corn-notill, Grass-Pasture, Woods, Soybean-clean, Wheat, Corn-mintill, Hay-windrowed, Soybean-mintill, Corn, Grass-Trees, Buildings-Grass-Trees-Drives, and Soybean-notill were considered in this study.

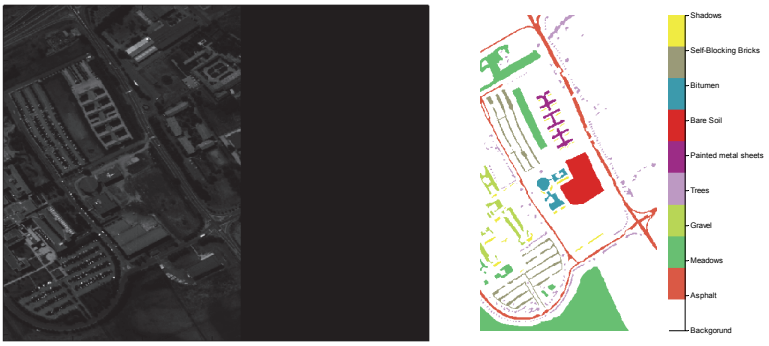

Fig. 3. The grayscale IR image of PAVIA with original size, 610×610 pixels and the ground truth of PAVIA with 610×340 pixels.

Table 2. Night categories and corresponding number of pixels in the Pavia University image. #(pixels) indicates the number of pixels per class.

No.	Category	#(pixels)	No.	Category	#(pixels)
1	Asphalt	6631	6	Bare Soil	5029
2	Meadows	18649	7	Bitumen	1330
3	Gravel	2099	8	Self-Blocking Bricks	3682
4	Trees	3064	9	Shadows	947
5	Painted metal sheets	1345			

The other hyeprspectral image used was PAVIS data set, a portion of the records from four flight lines over the urban area of Pavia in Northern Italy [15, 16]. It is a very fine-resolution hyperspectral data. The corresponding geometric resolution is 1.3 m. The flight was conducted in the framework of the HySens project, managed by Deutsches Zentrum fuer Luft-und Raumfahrt (the German AerospaceCenter) and sponsored by the

European Union within the Transnational Access to Major Research Infrastructures program (Contract no. HPRI-CT-1999-00075). This image was acquired by the Digital Airborne Imaging Spectrometer (DAIS). Its original size was 610×610 pixels with 103 spectral bands, but some of the samples in the image contained no information and had to be discarded before the analysis. Hence, the size used was reduced to 610×340 pixels. Nine classes were available in the original ground-truth image: Asphalt (Class 1), Meadows (Class 2), Gravel (Class 3), Trees (Class 4), Painted metal sheets (Class 5), Bare Soil (Class 6), Bitumen (Class 7), Self-Blocking Bricks (Class 8), and Shadows (Class 9). Fig. 3 shows the grayscale IR image and ground truth of PAVIA. Table 2 lists the number of pixels of each class.

Table 3. Descriptions of two UCI data sets.

Dataset	Classes	# of Samples	Features
Wine	3	178	13
WDBC	2	569	30

Two real UCI data sets, Wine and Breast Cancer Wisconsin (Diagnostic) (WDBC) were described in Table 3. The Wine dataset is a collection of data using three classes of wine from various locations in Italy. In the WDBC, there are two classes, benign and malignant. These data sets are available from the FTP server of the UCI [22] data repository.

4.2 Experimental Design

In the experiments, the multiclass SVM is based on one-against-all (OAA) multiclass strategy [18, 19]. In addition, the parameter of the RBF kernel of GDA, KNWFE, and SVM was all determined by APS [18]. However, the penalty parameter of SVM was still determined by CV in the same set, $\{2^{-5}, 2^{-4}, \ldots, 2^{5}\}$. For investigating the influence of training sample size to the dimensionality, three different cases, ill-posed, poorly posed, and well-posed cases, were considered for hyperspectral image classification. Let

$$N = N_1 + N_2 + \ldots + N_L.$$

In ill-posed case, $N_i = 10 < N < d$, where the data dimensionality exceeds the number of independent training samples in every class and the number of all training samples. In poorly-posed situation, $N_i = 40 < d < N$, where the dimensionality is still greater than the number of training samples in every class but the number of whole training samples is greater than the dimensionality. The third case, $d < N_i = 300 < N$ (well-posed case), there are enough independent training samples in every class. There are 100 testing samples per class. We randomly selected training samples and testing samples from data set to form 10 data subsets. Moreover, the average of testing accuracies through 10 subsets is used to compare the classification performance.

In addition, the comparisons between grid/automatic (G/A) version and traditional/theoretical (O/N) process of KNWFE and GDA were shown in the next section. Here, KNWFE (GDA) based on the grid search to find the best parameter with traditional process are denoted by KNWFE_GO (GDA_GO). KNWFE (GDA) based on the grid search

to find the best parameter with theoretical process are denoted by KNWFE_GN (GDA_GN). KNWFE (GDA) based on the automatic parameter selection method to find the best parameter with traditional process are denoted by KNWFE_AO (GDA_AO). KNWFE (GDA) based on the automatic parameter selection method to find the best parameter with theoretical process are denoted by KNWFE_AN (GDA_AN). Furthermore, the classification performance based on a feature selection method, F-Score [2], was also compared.

For the UCI data sets, about 10% (20%) of the samples were randomly chosen as training samples. All the samples were used as the testing set. This study employed overall classification accuracy (the percentage of the correctly classified samples for all classes) to measure the classification performance.

4.3 Experimental Results

Tables 4-6 show the average classification accuracies of SVM with specific GDA, KNWFE, and F-Score feature subsets and whole features of the IPS data set in case 1, 2, and 3, respectively. The classification accuracies by applying SVM with GDA or KNWFE features increase rapidly (compared to F-Score) in the first few feature subsets with dimensionalities, 1, 3, 5, and 10, among three cases. Moreover, KNWFE_AN and GDA_AN can determine more suitable feature subsets than KNWFE_AO and GDA_AO except that the classification performances are similar by applying GDA_AN and GDA_AO in case 3. In the ill-posed case (case 1), the highest accuracy (0.632) of KNWFE_AN+SVM is higher than those of KNWFE_AO+SVM, GDA_AN+SVM, GDA_AO+SVM, and F-Score+SVM. All of them are greater than the accuracy (0.570) of SVM with whole features. Moreover, only using feature subset of KNWFE_AN with 3 transformed features, the corresponding classification accuracy is similar to the accuracy of SVM with whole features.

In the poorly-posed case (case 2), the trends of accuracy of KNWFE_AN+SVM, KNWFE_AO+SVM, GDA_AN+SVM, and GDA_AO+SVM are similar to those in case 1. However, more features are required to have higher accuracies than the classification accuracy of SVM with whole features. In this case, the highest classification accuracies of KNWFE_AN+SVM, KNWFE_AO+SVM, GDA_AN+SVM, and GDA_AO+SVM are 0.789, 0.759, 0.777, and 0.761. All of them are still higher than the accuracy 0.750 obtain from SVM with whole features. Moreover, all of them are also higher than or similar to the highest accuracy 0.764 obtain from F-Score+SVM with 150 selected features.

Table 4. The average classification accuracies in Ill-posed case (Case 1) on IPS data set.

Method	Specific Dimensionality								
	1	3	5	10	20	30	50	120	220
KNWFE_AN	0.380	0.579	0.597	0.628	0.632	0.627	0.625	0.628	
KNWFE_AO	0.429	0.549	0.565	0.609	0.608	0.603	0.609	0.614	
GDA_AN	0.393	0.573	0.580	0.622	0.624	0.626	0.628	0.623	
GDA_AO	0.454	0.545	0.574	0.609	0.616	0.623	0.618	0.604	
F-Score	0.364	0.363	0.427	0.413	0.518	0.536	0.590	0.563	
SVM									0.570

Table 5. The average classification accuracies in poorly-posed case (Case 2) on IPS data set.

Method	Specific Dimensionality									
	1	3	5	10	20	30	50	100	150	220
KNWF_AN	0.396	0.651	0.720	0.782	0.789	0.785	0.785	0.785	0.780	0.781
KNWF_AO	0.470	0.615	0.680	0.732	0.739	0.738	0.736	0.746	0.753	0.759
GDA_AN	0.395	0.575	0.633	0.738	0.774	0.777	0.776	0.773	0.773	0.772
GDA_AO	0.467	0.621	0.694	0.738	0.761	0.755	0.743	0.723	0.716	0.710
F-Score	0.373	0.395	0.513	0.555	0.593	0.613	0.722	0.720	0.764	0.759
SVM										0.750

Table 6. The average classification accuracies in well-posed case (Case 3) on IPS data set.

Method	Specific Dimension									
	1	3	5	10	20	30	50	100	150	220
KNWF_AN	0.504	0.727	0.811	0.911	0.908	0.910	0.910	0.907	0.902	0.900
KNWF_AO	0.566	0.703	0.786	0.863	0.889	0.900	0.898	0.895	0.900	0.908
GDA_AN	0.600	0.717	0.797	0.890	0.900	0.902	0.899	0.898	0.894	0.894
GDA_AO	0.416	0.683	0.797	0.873	0.906	0.909	0.894	0.878	0.867	0.863
F-score	0.463	0.466	0.527	0.636	0.662	0.688	0.779	0.867	0.901	0.907
SVM										0.905

Table 7. The average classification accuracies in Ill-posed case (Case 1) on PAVIA data set.

Method	Specific Dimension								
	1	3	5	10	20	30	50	90	103
KNWF_AN	0.592	0.711	0.731	0.784	0.793	0.796	0.797	0.796	
KNWF_AO	0.633	0.713	0.719	0.744	0.769	0.778	0.780	0.779	
GDA_AN	0.603	0.744	0.758	0.765	0.770	0.775	0.780	0.788	
GDA_AO	0.611	0.704	0.717	0.752	0.771	0.770	0.756	0.757	
F-score	0.526	0.490	0.469	0.550	0.650	0.692	0.711	0.736	
SVM									0.733

Table 8. The average classification accuracies in poorly-posed case (Case 2) on PAVIA set.

Method	Specific Dimension							
	1	3	5	10	20	30	50	103
KNWF_AN	0.630	0.759	0.790	0.852	0.847	0.842	0.873	0.872
KNWF_AO	0.629	0.811	0.804	0.850	0.834	0.838	0.829	0.822
GDA_AN	0.629	0.773	0.796	0.858	0.862	0.859	0.860	0.828
GDA_AO	0.560	0.769	0.774	0.740	0.812	0.811	0.806	0.809
F-score	0.512	0.523	0.535	0.563	0.681	0.750	0.840	0.869
SVM								0.863

In case 3, the well-posed case, SVM with whole features has a good classification performance. Although the highest accuracy of GDA_AN+SVM is only 0.902, the highest accuracies of KNWFE_AN+SVM, KNWFE_AO, and GDA_AO are 0.911, 0.908, and 0.909 which are a little bit higher than the classification accuracy 0.905 by applying SVM with whole features.

Tables 7-9 illustrate the average classification accuracies of SVM with specific GDA, KNWFE, F-Score feature subsets and whole features of the PAVIA data set in

Table 9. The average classification accuracies in poorly-posed case (Case 3) on PAVIA set.

Method	Specific Dimension							
	1	3	5	10	20	30	50	103
KNWF_AN	0.483	0.773	0.854	0.917	0.933	0.942	0.944	0.947
KNWF_AO	0.629	0.851	0.888	0.921	0.922	0.919	0.927	0.926
GDA_AN	0.598	0.809	0.874	0.933	0.939	0.940	0.933	0.929
GDA_AO	0.491	0.787	0.881	0.932	0.914	0.912	0.921	0.918
F-score	0.516	0.544	0.543	0.553	0.716	0.778	0.869	0.910
SVM								0.914

cases 1-3, respectively. The results are quite similar to the results of IPS data set. In case 1, the highest accuracies of KNWFE_AN+SVM, KNWFE_AO+SVM, GDA+AN+SVM, GDA+AO+SVM, and F-Score are 0.797, 0.780, 0.788, 0.771, and 0.736 respectively. In addition, KNWFE_AN and GDA_AN outperforms than KNWFE_AO and GDA_AO.

In case 2, the highest accuracies of KNWFE_AN+SVM, KNWFE_AO+SVM, GDA+AN+SVM, GDA_AO+SVM, and F-Score are 0.873, 0.850, 0.862, 0.812 and 0.869, respectively. But the classification accuracies of GDA_AO are poorer than the classification (0.863) of SVM. However, the highest classification accuracy of GDA_AN are similar to the classification of SVM. It shows that the theoretical process can improve the classification performance. In case 3, they are 0.947, 0.927, 0.940, 0.932 and 0.910, respectively. All of them are higher than the classification performance of SVM.

From Tables 4-9, we can find that if only about half features of KNWFE_AN and GDA_AN are selected, the corresponding classification accuracies are similar to the classification accuracy of SVM with whole features among all cases on both hyperspectral image data sets. Hence applying KNWFE_AN or GDA_AN can obtain good combinations (for classification) of original features. Moreover, applying KNWFE_AN and GDA_AN can improve the classification a lot in the ill-posed case (case 1). Someone may find the numbers of largest feature subset of KNWFE_AN and GDA_AN in case 1 on IPS and PAVIA data sets are 120 and 90, respectively. This is due to the size of kernel matrix with respect to the whole training samples. For the IPS data set, there are N=120 training sample in case 1, and there are N=90 training sample in case 1 for the PAVIA data set.

Furthermore, from Tables 6 and 9, the classification accuracies, 0.911 and 0.917, of KNWFE_AN based on 10 selected features are close to the classification accuracies 0.905 and 0.914 of SVM due to enough independent training samples in every class. Hence, SVM without whole features can get an acceptable classification performance in the well-posed case. However, in the ill-posed and poorly-posed cases, GDA or KNWFE are suggested to reduce the dimensionality and then the classification performance can be improved.

Table 10 shows the highest classification accuracies of KNWFE_AN+SVM, KNWFE_AO+SVM, KNWFE_GN+SVM, KNWFE_GO+SVM, GDA_AN+SVM, GDA_AO+SVM, GDA_GN+SVM, GDA_GO+SVM, F-Score, and the classification accuracy of SVM with whole features in two UCI data sets, Wine and WDBC, with 10% and 20% randomly selected samples as training samples. Moreover, the computation time for determining the best parameters of RBF kernel were compared, too. Using automatic parameter selection method to find the best parameter is more effective than using grid

search to find the best parameter. In both UCI data set, the highest classification accuracies of KNWFE_AN+SVM, KNWFE_GN+SVM, GDA_AN+SVM and GDA_GN+SVM are higher than the classification accuracy of SVM with whole features. But some highest classification accuracies determined by traditional eigenvalue resolution are lower than the classification accuracy of SVM with whole feature.

Table 10. The highest classification accuracies with the corresponding numbers of selected features (in brackets) of KNWFE_AN+SVM, KNWFE_AO+SVM, KNWFE_GN+SVM, KNWFE_GO+SVM, GDA_AN+SVM, GDA_AO+SVM, GDA_GN+SVM, GDA_GO+SVM, F-Score and SVM with the whole features in two UCI data sets.

Dataset	Method	10%		20%	
		Accuracy	Time (s)	Accuracy	Time (s)
Wine	KNWFE_AN	0.807 (2)	0.013	0.924 (12)	0.009
	KNWFE_AO	0.681 (6)	0.011	0.728 (5)	0.013
	KNWFE_GN	0.788 (4)	7.135	0.857 (3)	6.903
	KNWFE_GO	0.679 (5)	7.183	0.735 (4)	6.889
	GDA_AN	0.730 (2)	0.014	0.833 (2)	0.011
	GDA_AO	0.690 (5)	0.012	0.748 (8)	0.009
	GDA_GN	0.748 (12)	7.114	0.808 (2)	6.653
	GDA_GO	0.696 (6)	7.124	0.771 (11)	6.791
	F-Score	0.731 (3)		0.758 (1)	
	N/A	0.725 (13)		0.738 (13)	
WDBC	KNWFE_AN	0.947 (20)	0.014	0.968 (12)	0.021
	KNWFE_AO	0.922 (11)	0.012	0.951 (15)	0.023
	KNWFE_GN	0.909 (10)	2.623	0.946 (30)	2.935
	KNWFE_GO	0.830 (14)	2.698	0.933 (3)	3.011
	GDA_AN	0.921 (1)	0.014	0.953 (11)	0.022
	GDA_AO	0.902 (25)	0.012	0.935 (8)	0.022
	GDA_GN	0.928 (17)	2.662	0.933 (18)	2.782
	GDA_GO	0.825 (21)	2.573	0.912 (22)	2.890
	F-Score	0.914 (3)		0.923 (21)	
	N/A	0.915 (30)		0.925 (30)	

5. CONCLUSION

Although, in the previous study [6], the classification performance with KNWFE feature subset outperforms that with original features (similar results in this study), the RBF kernel parameter with KNWFE was determined by try and error. Hence, we cannot guarantee the corresponding classification performance is good enough. Moreover, the implicit transformation matrix of GDA and KNWFE was not so accurate due to the solution procedure of the generalized eigenvalue problem. In this study, a theoretical procedure for solving the generalized eigenvalue problem directly based on implicitly kernel-ased within- and between-class scatter matrices was proposed.

Moreover, APS [12, 13] was applied to pre-determine automatically the RBF kernel parameter of GDA, KNWFE, and SVM in few seconds. Hence, a fast and automatic kernel-based classification scheme, kernel-based feature extraction method (GDA or

KNWFE) with kernel-based classifier (SVM), was proposed. From the experimental results on two hyperspectral image data sets and two UCI data sets, SVM with appropriate RBF kernel parameter can have an almost perfect classification performance. Additionally, in the small sample size problem, the proposed faster and automatic kernel-based classification scheme, KNWFE_AN+SVM and GDA_AN+SVM, can archive higher classification accuracy based on few selected features. Now, two main factors that influence the classification performance by applying kernel-based feature extraction method have been solved in this study. Therefore, in the future, the readership can manipulate this suggest kernel-based classification scheme more accurately and in an easy way.

ACKNOWLEDGMENT

The authors would like to thank the participants for this study. They are also grateful to Dr. Landgrebe for providing the Indian Pine site data set and Prof. Paolo Gamba and Prof. Fabio Dell'Acqua for the PAVIA data set. This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 103-2221-E-142-03-MY3 and 102-2218-E-142-001-, and by the Natural Science Foundation of Fujian Provincial (No. 2017J01780).

REFERENCES

1. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, San Diego, CA, 1990.
2. I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, *Feature Extraction*, *Foundations and Applications*, Springer, Berlin Heidelberg New York, 2006.
3. X. P. Jia, B. C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proceedings of the IEEE*, Vol. 101, 2013, pp. 676-697.
4. G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, Vol. 3, 2006, pp. 93-97.
5. G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, Vol. IT-14, 1968, pp. 55-63.
6. B. C. Kuo, C. H. Li, and J. M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, 2009, pp. 1139-1155.
7. B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, 2004, pp. 1096-1105.
8. G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, Vol. 12, 2000, pp. 2385-2404.
9. G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*, John Wiley & Sons, Ltd., 2009.
10. D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, NJ, 2003.

11. J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, UK, 2004.
12. C. H. Li, C. T. Lin, B. C. Kuo, and H. S. Chu, "An automatic method for selecting the parameter of the RBF kernel function to support vector machines," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 2010, pp. 836-839.
13. C. H. Li, H. H. Ho, Y. L. Liu, C. T. Lin, B. C. Kuo, and J. S. Taur, "An automatic method for selecting the parameter of the normalized kernel function to support vector machines," *Journal of Information Science and Engineering*, Vol. 28, 2012, pp. 1-15.
14. C. H. Li, B. C. Kuo, L. H. Lin, W. Wu, and D. Lan, "Apply an automatic parameter selection method to generalized discriminant analysis with RBF kernel for hyperspectral image classification," in *Proceedings of International Conference on Machine Learning and Cybernetics*, 2013, pp. 253-258.
15. F. Dell'Acqua, P. Gamba, and A. Ferrari, "Exploiting spectral and spatial information for classifying hyperspectral data in urban areas," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 2003, pp. 464-466.
16. J. A. Benediktsson, A. J. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, 2005, pp. 229-243.
17. B. C. Kuo, H. H. Ho, C. H. Li, C. C. Hung, and J. S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7, 2014, pp. 317-326.
18. C. H. Li, B. C. Kuo, C. T. Lin, and C. S. Huang, "A spatial-contextual support vector machine for remotely sensed image classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, 2012, pp. 784-799.
19. C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, Vol. 13, 2002, pp. 415-425.
20. C. E. Brodley and P. Smyth, "Applying classification algorithms in practice," *Statistics and Computing*, Vol. 7, 1997, pp. 45-56.
21. É. Perthame, C. Friguet, and D. Causeur, "Stability of feature selection in classification issues for high-dimensional correlated data," *Statistics and Computing*, 2015.
22. C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, http://www.ics.uci.edu/~mlearnMLRepository.html, 1998.

Cheng-Hsuan Li (李政軒) received the B.S. and M.S. degrees in Department of Applied Mathematics from National Chung Hsing University, Taichung, Taiwan, in 2001 and 2003, respectively, and the Ph.D. degree in Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 2012. He is currently an Associate Professor with the Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taiwan. His research interests in-

clude pattern recognition, statistical learning, machine learning, and nonparametric cognitive diagnosis model.

Pei-Jyun Hsieh (謝佩鈞) received the MS degrees from National Taichung University of Education, Taichung, Taiwan, in 2015, respectively. She is currently working toward the Ph.D. degree in Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung, Taiwan. She is currently an Executive Assistant with the Graduate Institute of Educational Information and Measurement, National Taichung University, Taichung. Her research interests include machine learning.

Li-Hui Lin (林丽惠) received the B.S. degree in College of Mathematics and Computer Science from Fujian Normal University, Fujian, China, in 2002, and the Master's degree of engineering in College of Mathematics and Computer Sciences from Fuzhou University, Fujian, China, in 2007. She is currently an Associate Professor with College of Mathematics and Computer Science of Wuyi University. Her research interests include pattern recognition, statistical learning, machine learning.