Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **2**, March **2016**

A Unified Framework of Multiple Kernels Learning for Hyperspectral Remote Sensing Big Data

Xiao-Dan Xie, Bo-Hu Li
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
xiexiaodanbeijing@126.com

Xu-Dong Chai
Beijing Simulation Center
Beijing, China

Received August, 2015; revised September, 2015

Abstract. *Analysis on hyperspectral remote sensing big data is widely in remotely sensed satellite imaging and aerial reconnaissance, the development of sensor technology brought the developing of collecting image data using hyperspectral instruments with hundreds of contiguous spectral channels. Machine learning based hyperspectral sensing data analysis is a feasible way, and among these machine learning methods, kernel learning is a feasible nonlinear feature extraction on hyperspectral sensing data. This paper is to solve the problem of the nonlinear kernel function selection, to improve the system performances of recognition and prediction accuracy. A framework of multiple kernel learning is proposed for classification on hyperspectral remote sensing big data, and some experiments are implemented on two hyperspectral image databases. The comprehensive experiments show that the proposed algorithm is effective on hyperspectral remote sensing big data.*
Keywords: Hyperspectral Remote Sensing; Big Data; Kernel Learning; Multi-kernel Learning Classification; Image Classification.

1. **Introduction.** The development of sensor technology brought the developing of collecting image data using hyperspectral instruments with hundreds of contiguous spectral channels. Hyperspectral imagery is the most popular remote sensing technology on satellite platform, with the prospective applications in military monitoring, energy exploration, geographic information, and so on. Hyperspectral instruments with hundreds of contiguous spectral channels brings the developing of collecting remote imagery data. The increasing spectral and space resolution bring a large size of data, which brings two problems in the practical applications: 1) the bandwidth of the communication channel limits the transmission of the full hyperspectral image data for the further processing and analysis on the ground; 2) the demand of the real-time processing for some applications. Data compression is a solution to the transmission problem but no ability for the real-time analysis. So, machine learning-based data analysis technology is feasible and effective to produce one image from the full band of hyperspectral images through classifying the spectrum curve of each pixel according to the spectrum data of each object. For the hyperspectral sensing data classification, we present a multikernel machine framework for hyperspectral remote imagery system. Motivated by the fact that kernel machine is effective to the nonlinear classification but the performance of kernel-based system is

largely influenced by the function and parameter of kernel, we present a framework of multiple kernels learning. Dimensionality reduction method is a most popular method for feature extraction, and many dimensionality reduction methods are proposed in the previous works such as Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) [1]. LDA works well in some cases but fails to capture a nonlinear relationship with a linear mapping. In order to solve the nonlinear problems, kernel method is used to represent the complicated nonlinear relationships of input data. Kernel version of linear dimensionality reduction methods are developed in recent years, such as Kernel PCA (KPCA), Kernel Discriminant Analysis (KDA) [17]. In the following research, many linear learning algorithms are kernelized to develop the novel kernel learning methods. KDA has been applied in many real applications owing to its excellent performance on feature extraction. Researchers have developed a series of KDA algorithms (Baudat and Anouar [2], Liang and Shi [12], Wang [15], Chen [4] and Lu [13]). Moreover, in recent research Locality Preserving Projection (LPP) has been widely studied and used in many areas as the important manifold learning. Researchers presented an alternative formulation of Kernel LPP (KLPP) to develop a framework of KPCA+LPP algorithm in the previous work [10, 7] for face recognition and radar target recognition, and other researchers improved LPP with kernels in the previous works [19, 6, 18] and [8].

On current kernel learning methods, the performance of many linear learning methods is improved because the data distribution in the nonlinear feature space is easy to classification owing to kernel mapping. The geometrical structure of the data in the kernel mapping space, which is totally determined by the kernel function, has significant impact on the performance of these kernel learning methods. The discriminative ability of the data in the feature space could be even worse if an inappropriate kernel is used. In the previous work, researchers optimized the parameters of kernel function to improve KDA [8, 14, 5], but these methods only choosing the optimal parameter of kernel from a set of discrete values which are created in advance. The geometry structure of data distribution in the kernel space is not be changed only through the changing the parameters of kernel. Xiong proposed a data-depend kernel for kernel optimization [16], and Amari presented support vector machine classifier through modifying the kernel function [1]. In the previous works [14, 9], authors present data-dependent kernel based KDA algorithm for face recognition application. Moreover, multiple kernel learning methods are developed, for example, Sparse Multiple Kernel Learning [21], Large Scale Multiple Kernel Learning[22], lp-Norm Multiple Kernel Learning[23].And kernel learning method is applied to hyperspectral remote sensing imagery classification [24], Multiple kernel extreme learning machine.[25], and multimodal analysis on Alzheimer's disease using multiple kernel learning [26], and the feasibility and excellent performance were reported in this work.

As above discussion, kernel learning is an important research topic in the machine learning area, and some theory and applications fruits are achieved and widely applied in pattern recognition, data mining, computer vision, image and signal processing areas. The nonlinear problems are solved with kernel function, and system performances such as recognition accuracy, prediction accuracy are largely increased. However, kernel learning method still endures a key problem, i.e., kernel function and its parameter selection. Kernel function and its parameters have the direct influence on the data distribution in the nonlinear feature space, and the inappropriate selection will influence the performance of kernel learning.

2. **Algorithm.** Kernel learning has its first key step of calculating the kernel matrix with the selected kernel function for the classification, clustering and other statistical pattern analysis. Two problems occur during this step. One is the heavy computing occurs

for the computing the kernel matrix using all the training samples. Second is the high performance influence problem of kernel function and its parameter owing the geometry structure of sample data is different with the different kernel function mapping. Kernel Discriminant Analysis (KDA) and Kernel Principal Component Analysis (KPCA) are analyzed as follows.

KDA transforms the transformation matrix from the input space to a nonlinear high-dimensional feature space [22]. Given L classes of M training samples $\{x_1, x_2, ..., x_M\}$ in an N-dimensional space R^N, the data are mapped into a feature space F via the following nonlinear mapping:

$$\Phi : R^N \to F, xa\ \Phi(x) \tag{1}$$

The Fisher criterion in the feature space F is defined by

$$J(V) = \frac{V^T S_B^{\Phi} V}{V^T S_T^{\Phi} V} \tag{2}$$

where V is the discriminant vector, and S_B^{Φ} and S_T^{Φ} are the between-class scatter matrix and total-scatter matrix, respectively. Any solution V belongs to the span of all training patterns in R^N. Hence, there exists coefficients $c_p(p = 1, 2, ..., M)$ such that

$$V = \sum_{p=1}^{M} c_p \Phi(x_p) = \Psi\alpha \tag{3}$$

where $\Phi = [\Phi(x_1), \Phi(x_2), ..., \Phi(x_M)]$ and $\alpha = [c_1, c_2, ..., c_M]^T$. Assuming that the data are centered, the Fisher criterion is transformed into

$$J(\alpha) = \frac{\alpha^T KGK\alpha}{\alpha^T KK\alpha} \tag{4}$$

where $G = diag(G_1, G_2, ..., G_L)$, G_i is an $n_i \times n_i$ matrix whose elements are $\frac{1}{n_i}$, and the kernel matrix K is calculated by a basic kernel $k(x, y)$.

KPCA is to project the input data from the linear space into the nonlinear space, and then implement PCA in the nonlinear feature space for feature extraction[13]. For the clear description, we firstly introduce PCA and then kernelize PCA into KPCA. Given the training samples $x_1, x_2, ...x_n$ then $C = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$, where $\bar{x}$ is the mean sample of all training samples, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Generally, these eigenvectors are calculated in many practical applications, SVD (Singular Value Decomposition) is applied into solving the singular matrix problem. Let $Q = [x_1 - \bar{x}, ...x_n - \bar{x}]$, then $C = \frac{1}{n}QQ^T$, then $R = Q^T Q$ is the $n \times n$ positive definite matrix. The dimension of R is less than the dimension of C . According to SVD, the eigenvectors $v_1, v_2, ..., v_m$ according to the m largest values $(\lambda_1 \geq \lambda_1 \geq ... \geq \lambda_m)$,the projection vectors are computed via $w_j = \frac{1}{\sqrt{\lambda_j}}Qv_j$, $j = 1, 2, ..., m$ .For any sample x ,the jth feature is $y_j = w_j^T x = \frac{1}{\sqrt{\lambda_j}}v_j^T Q_x^T, j = 1, 2, ..., m$. PCA is kernelized as follows. $C = \frac{1}{n}\sum_{i=1}^{n}(\Phi(x_i) - \bar{\Phi})(\Phi(x_i) - \bar{\Phi})^T$, where $\bar{\Phi} = \frac{1}{n}\sum_{i=1}^{n}\Phi(x_i)$, and let $C' = \frac{1}{n}\sum_{i=1}^{n}(\Phi(x_i))(\Phi(x_i))^T$ and $Q = [\Phi(x_1), \Phi(x_2), ...\Phi(x_n)]$, then $C' = \frac{1}{n}QQ^T$. According to $R' = Q^T Q$, with the kernel function. Compute the eigenvectors $u_1, u_2, ...u_m$ according to the mth eigenvalue $(\lambda_1 \geq \lambda_1 \geq ... \geq \lambda_m)$ of R, then $w_1, w_2, ..., w_m$ is calculated by

$$w_j = \frac{1}{\sqrt{\lambda_j}}Qu_j, j = 1, 2, ..., m \tag{5}$$

Then

$$y_j = w_j^T x = \frac{1}{\sqrt{\lambda_j}} u_j^T [k(x_1, x), k(x_2, x), ..., k(x_n, x)] \tag{6}$$

Kernel mapping has its first key step of calculating the kernel matrix with the selected kernel function for the classification, clustering and other statistical pattern analysis.

The defintion of Multiple Kernel Learning (MKL) can be defined as

$$k(x, x') = \sum_{i=1}^{n} d_i k_{0,i}(x, x') \tag{7}$$

where $x, x' \in \mathbb{R}$, $k_{0,i}(x, x')$ is the ith basic kernel, n is the number of basic kernels for combination.

We determine the weights d_i , $i = 1, 2, ..., I$ through solving the convex optimization as

$$\begin{aligned} & max\ \Psi(K_0, K^*) \\ & s.t.\ K_0 = \sum_{i=1}^{I} d_i K_{0,i}, tr(K_0) = 1, d_i \geq 0, \forall i \end{aligned} \tag{8}$$

where $K^*(x, x^{'})$ is the ideal target kernel, tr is the trace of a matrix, $\Phi(K_0, K^*) = \langle K_0, K^* \rangle_F / \|K^*\|_F \|K_0\|_F$, $\langle ., . \rangle$ is the Frobenius norm between two matrices, and $\|K^*\|_F$ is a constant corresponding to a certain classification task. $H = [v(K_{0,1})\ v(K_{0,2})\ ...\ v(K_{0,I})]$,$v(.)$ is defined as

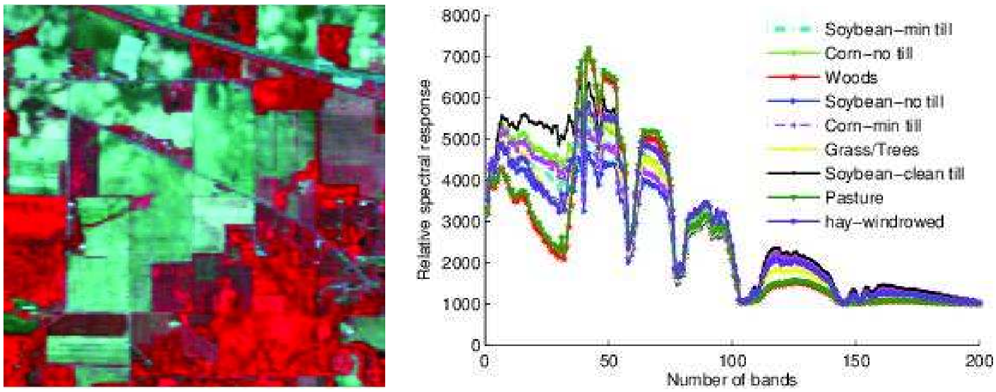
$$\begin{aligned} & min - v(K^*)^T H d + \gamma_1 \|d\|_2^2 \\ & s.t.\ d^T H^T H d \leq 1, d_i \geq 0, \forall i \end{aligned} \tag{9}$$

where $d = [d_1\ d_2\ ...\ d_I]^T$ is weight to be solved, $\|d\|_2^2$ is the regularization, and $\gamma_1 \in [0, 1]$ is the regularization parameter. The matrix H is the constant value of matrix computed by kernel vector, and the matrix is constructed by the training samples, and the matrix represents the relationship between the different training samples. The different matrix brings the different solution for the practical application. So the performance of data-driven kernel learning is determined by the matrix H. When the matrix is determined, and the solution is also be solved for the practical application.

3. **Experiments and discussion.**

3.1. **Experiments Setting and databases.** In the experiments, we evluate the performance of multiple kernels learning based SVM method on Hyperspectral image data classification. The experiments are implemented on a Pentium 3.0 GHz computer with 512MB RAM and programmed in MATLAB and the procedural parameters are chosen with cross-validation method. on Indian Pines data, which has various spectral and spatial resolutions reflecting different environments of remote sensing are adopted in the experiments. The first test set to be used was the well-known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image scene. The data have the spectral resolution of 224 bands covering the 0.4-2.5 range and spatial resolution of 20m per pixel. In the experiments, we remove the noisy and water-vapor absorption bands, 200 bands reserved in the experiments. The hyperspectral cube has the whole 145 145 pixels scene, and it has 16 classes of objects, ranged with the size from 20 to 2468 pixels. In the experiments, we applied only 9 classes of training samples, and some examples are shown in Figure.1.

In the experiments, we choose the procedural parameters through cross-validation method for the practical application. All training samples are considered the samples



(a) (b)

FIGURE 1. Indian Pines data. (a) Three band false color composite. (b) Spectral signatures.

to cross-validation method. In the experiments, we crope the original image to a size of 211×307, which composed of 7 classes of land-covers including roof, grass, street, trees, water, path and shadow. Some examples are shown in Figure 2. The training set includes the seven training samples set. The training samples are chosen randomly from all pixels, and the rest pixels are as the test samples set.

3.2. **Results.** On the hyperspectral image database, Indian Pines data, we have the averaged accuracy of 10 times of experiments as the final result. We implement Support Vector Classifier (SVC), Kernel Sparse Representation Classifier (KSRC) to hyperspectral image classification. We test multikernels for kernel classifiers on SVC and KSRC, that is, PK-SVC: Polynomial Kernel-SVC, GK-SVC: Gaussian Kernel-SVC, MK-SVC: Multi-kernels Based SVC, PK- KSRC: Polynomial Kernel- KSRC, GK-KSRC: Gaussian Kernel-KSRC, MK- KSRC: Multi-kernels Based KSRC. For the quantitative comparison, we implement some experiments using the averaged accuracy to evaluate the performance of the algorithms, including PK-SVC, GK-SVC, MK-SVC, PK- KSRC, GK-KSRC, MK-KSRC. The experimental results are shown in Table 1 and Table 2. On the SVC, MK-SVC performs better than PK-SVC and GK-SVC. On the KSRC, MK-KSRC outperforms PK-KSRC and GK-KSRC. Especially, the polynomial kernel performs better than Gaussian kernel under SVC and KSRC classifiers. On the multiple kernels, Gaussian kernel and Polynomial kernel are as the basic kernels to combination of multiple kernels.

TABLE 1. Performance of SVC on the Indian Pines data (%)

Class	1	2	3	4	5	6	7	8	9	10	11	12
PK-SVC	49.3	58.7	96.4	39.2	65.8	93.6	62.9	85.3	100	65.8	72.3	58.4
GK-SVC	78.0	73.6	99.1	76.9	80.5	97.1	79.7	89.8	99.7	83.6	86.0	80.7
MK-SVC	78.2	74.8	99.3	78.9	86.5	98.2	81.4	95.2	99.8	85.2	87.1	83.5

The results are shown in Table 1 and Table 2, multiple kernel learning is effective and feasible to improve the recognition system on hyperspectral sensing data classification system. Motivated by the fact that kernel machine is effective to the nonlinear classification but the performance of kernel-based system is largely influenced by the function and parameter of kernel, optimizing the parameters not effective to promote the kernel-based learning system owing to the unchanged data structure with the changing of the parameter of kernel function. No a universal single kernel is very effective way to detecting

Table 2. Performance of KSRC on the Indian Pines data (%)

Class	1	2	3	4	5	6	7	8	9	10	11	12
PK-KSRC	51.8	59.6	96.1	49.1	78.5	93.8	62.8	84.7	100	67.5	75.2	60.7
GK-KSRC	77.8	76.4	99.1	75.5	79.0	97.4	82.7	88.7	100	83.9	86.3	81.1
MK-KSRC	78.2	82.3	99.5	81.4	90.4	98.2	82.4	96.4	100	85.3	88.3	83.5

intrinsic information for the complicate sample data in the input data space. Multiple kernels are combined to more precisely characterize the data for improving performance on solving complex visual learning tasks. Data compression is a solution to the transmission problem but no ability for the real-time analysis. So, machine learning-based data analysis technology is feasible and effective to produce one image from the full band of hyperspectral images through classifying the spectrum curve of each pixel according to the spectrum data of each object. The hyperspectral data machine learning system is implemented on the satellite platform. After the hyperspectral data collection, each pixel is classified and denoted to the different objects based on the spectrum database. The spectrum data in database is collected in advance, so it has inconsistency between the spectrum with the data collection. The inconsistency can be consider the nonlinear changing. The relationships of between spectral curves is the classical nonlinear relationship. So the classification is the nonlinear and complex classification problem. Traditional classification methods are not effective to hyperspectral sensing data, among these machine learning methods kernel learning is a feasible and effective nonlinear classifier methods on hyperspectral sensing data.

From the above results, we can conclude as follows.

Because the experimental limitation, we only have the comprehensive analysis on less memory resources.Kernel-based image classification system can process the image matrix, and the currently kernel functions used in kernel learning method are vector-based function. All images must be transformed to vectors, and these vectors must be saved for the kernel-based learning system. And in the thus image classification system, the original image and transformed vector must be saved for the kernel learning. The input of the kernel function is vector or an $[N \times 1]$ matrix. For an $[m \times n]$ matrix, the matrix must be transformed to a vector of $[M \times 1]$, where $M = m \times n$. Thus, the one image matrix and vector must be saved, that is, for a $[m \times n]$ pixels of image, the save space is $m \times n \times 2$ of pixels for the traditional kernel learning, but only $m \times n$ of pixels of saving space for the matrix-kernel learning. So, for the proposed matrix-kernel learning will save 50% of memory resources.

On the experimental the selection of the algorithms' parameters, we have some detailed experiments as follows. In the experiments, we choose the procedural parameters through cross-validation method for the practical application. For the detail description, the training dataset is randomly divided to training sub-dataset and test sub-dataset, and the parameters is to train the classifier with training sub-dataset then the test sub-dataset is to test the performance of the parameters. The kernel functions, Gaussian kernel and polynomial kernel, the parameters determined through cross-validation method. Moreover, the free parameter is chosen through cross-validation method.

4. **Conclusions.** This paper presents a framework of multiple kernels learning method, and this framework is to solve the selection of function and parameter of kernel, which have the heavy influences on the performance of kernel-based learning system. Multiple kernels are combined to more precisely characterize the data for improving performance on solving complex visual learning tasks. The performance of the framework is testified

using the many databases, and the learning framework is feasible to the hyperspectral image classification. The framework can be used to other kernel-based systems in the practical applications.

REFERENCES

[1] S. I. Amari and S. Wu, Improving support vector machine classifiers by modifying kernel functions,*Neural Networks* , vol.12, no.6, pp.783-789, 1999.

[2] G. Baudat and F. Anouar, Generalized discriminant analysis using a kernel approach,*Neural computation*,vol.12, no.10, pp.2385-2404, 2000.

[3] Belhumeur, N. Peter and Hespanha, João P and Kriegman, David J,Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,*IEEE Transactions. Pattern Analysis and Machine Intelligence,* , vol.19, no.7,pp.711-720, 1997.

[4] W. S. Chen, P. C. Yuen, J. Huang and D. Q. Dai, Kernel machine-based one-parameter regularized Fisher discriminant method for face recognition,*IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.35, no.4, pp.659-669, 2005.

[5] M. N. Kan, D. Xu, S. Q. Shan and X. L. Chen, Semisupervised hashing via kernel hyperplane learning for scalable image search,*IEEE Trans. Circuits and Systems for Video Technology*, vol.24, no.4, pp.704-713, 2014.

[6] J. Cheng, Q. S. Liu, H. Q. Lu and Y. W. Chen, Supervised kernel locality preserving projections for face recognition,*Neurocomputing*, vol.67, pp.443-449, 2005.

[7] G. V. Feng, D. W. Hu, D. Zhang and Z. T. Zhou, An alternative formulation of kernel LPP with application to image recognition,*Neurocomputing*, vol.69, no.13, pp.1733-1738, 2006.

[8] J. Huang, P. C. Yuen, W. S. Chen and J. H. Lai, Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition,*IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth*, pp.327-332, 2004.

[9] J. B. Li and J. S. Pan and Z. M. Lu, Kernel optimization-based discriminant analysis for face recognition,*Neural Computing and Applications*,vol. 18, no. 6, pp. 603-612, 2009.

[10] J. B. Li, J. S. Pan and S. C. Chu, Kernel class-wise locality preserving projection,*Information Sciences*, Vol. 178, No. 7, pp. 1825-1835, 2008.

[11] J. B. Li, L. J. Yu and S. H. Sun, Refined kernel principal component analysis based feature extraction,*Chinese Journal of Electronics*, Vol.20, No. 3, pp. 467-470, 2011.

[12] Z. Z. Liang and P. F. Shi, Uncorrelated discriminant vectors using a kernel method,*Pattern Recognition*, Vol.38, No. 2, pp. 307-310, 2005.

[13] J. W. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms,*IEEE Trans. on Neural Networks*, Vol. 14, No. 1, pp.117-126, 2003.

[14] J. S. Pan, J. B. Li and Z. M. Lu, Adaptive quasiconformal kernel discriminant analysis,*Neurocomputing* , vol.71 ,no.13 , pp.2754–2760, 2008.

[15] L. Wang, K. L. Chan and P. Xue, A criterion for optimizing kernel parameters in KBDA for image retrieval,*IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 35, No. 3, pp. 556–562, 2005.

[16] H. L. Xiong , Swamy MNS and Ahmad M Omair, Optimizing the kernel in the empirical feature space,*IEEE Trans. Neural Networks*, vol.16, NO.2, pp.460–474,2005.

[17] M. H. Yang, Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods, pp.0215, 2002.

[18] H. T. Zhao, S. V. Sun, Z. L. Jing and J. Y. Yang, Local structure based supervised feature extraction,*Pattern Recognition*, Vol.39, no.8, pp. 1546–1550,2006.

[19] Q. Zhu, Reformative nonlinear feature extraction using kernel MSE,*Neurocomputing*, vol. 73, no.16, pp. 3334–3337,2010.

[20] T. Devis, C.-V. Gustavo, M. Giona and K. Mikhail, Learning relevant image features with multiple-kernel classification,*IEEE Trans. Geoscience and Remote Sensing*, Vol.48, no.10, pp. 3780–3791, 2010.

[21] S. Niranjan and Y. C. Shin, Sparse multiple kernel learning for signal processing applications,*IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp.788–798, 2010.

[22] Sonnenburg Sören, Rätsch Gunnar, Schäfer Christin and Schölkopf, Bernhard,Large scale multiple kernel learning,*The Journal of Machine Learning Research*, vol.7, pp. 1531–1565, 2006.

[23] M. Kloft, U. Brefeld, Sonnenburg Sören and Zien Alexander, Lp-norm multiple kernel learning,*The Journal of Machine Learning Research*, vol. 12, pp. 953–997, 2011.

[24] Chen Chen, Li Wei, Su Hongjun, Liu Kui, Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine,*Remote Sensing*, vol. 6, no. 6, pp. 5795–5814, 2014.

[25] X. W. Liu, L. Wang, G. B. Huang, J. Zhang and J. P. Yin, Multiple kernel extreme learning machine,*Neurocomputing* , vol. 149 , pp.253–264, 2015.

[26] M. Dyrba, M. Grothe, T. Kirste and S. J. Teipel, Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM,*Human brain mapping*, vol.36,no. 6, pp. 2118–2131, 2015.