View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 93 (2016) 396 – 402

Procedia Computer Science

6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Dimensionality Reduction using Band Selection Technique for Kernel based Hyperspectral Image Classification

Reshma.R[a,*], V.Sowmya[a], K.P.Soman[a]

[a]*Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India*

Abstract

Hyperspectral images have abundant of information stored in the various spectral bands ranging from visible to infrared region in the electromagnetic spectrum. High data volume of these images have to be reduced, preserving the original information, to ensure efficient processing. In this paper, dimensionality reduction is done on Indian Pines and Salinas-A datasets using inter band block correlation coefficient technique followed by Singular Value Decomposition (SVD) and QR decomposition. The dimensionally reduced images are classified using GURLS and LibSVM. Classification accuracies of the original image is compared to that of the dimensionally reduced image. The experimental analysis shows that, for 10% training sample the overall accuracy , average accuracy and kappa coefficient of the dimensionally reduced image (about 50% of the dimension is reduced) is i)83.52%, 77.18%, 0.8110 for Indian Pines and ii)99.53%, 99.40%, 0.9941 for Salinas-A dataset which is comparable to that of original image i)84.67%, 82.28%, 0.8247 for Indian Pines and ii)99.32%, 99.18%, 0.9916 for Salinas-A dataset.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the Organizing Committee of ICACC 2016

Keywords: Dimensionality reduction; hyperspectral image; classification; kernel methods; GURLS; LibSVM; accuracy; band selection.

1. Introduction

Hyperspectral imaging have a remarkable development in the recent years with the advancement in hyperspectral sensors. High resolution and abundance of information in the hyperspectral images have attracted researchers from various fields like agriculture, environmental monitoring, mining, etc. The volume of data which is acquired at any time (t) is very high which pose a problem in data storage and processing. This reveals the importance of reducing the dimension of these images.

Dimensionality reduction[1,2] is a process in which a higher dimensional data is represented using a small portion of the data ie, lower dimension , preserving its original features. Various techniques used for dimension reduction are Principal Component Analysis (PCA), Discriminant Analysis, Minimum noise fraction (MNF), etc[3]. Inter band block correlation coefficient followed by Singular Value Decomposition(SVD) and QR decomposition is utilized in

* Corresponding author. Tel.: +91-9790446238.
E-mail address: reshmaraju2992@gmail.com

1877-0509 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the Organizing Committee of ICACC 2016
doi:10.1016/j.procs.2016.07.226

the proposed work for reducing the dimension of the hyperspectral images[4].

Hyperspectral classification techniques are applied to the dimensionally reduced HSI. Orthogonal Matching Pursuit (OMP), Regularized Least Square (RLS), Support Vector Machine (SVM), Subspace Pursuit (SP), Relevance Vector Machine (RVM), etc are the commonly used methods for classification. In the experiment we conducted, classification is performed using kernel based libraries[5,6,7] like Grand Unified Regularized Least Squares (GURLS)[8] and LibSVM[9]. The classification accuracy of dimensionally reduced image and original image is compared using various accuracy assessment measures[7].

2. Dimensionality Reduction

Hyperspectral images have redundant information due to its strong spectral correlation. So, it is useful to reduce the dimension of these images. Dimensionality reduction can be done based on selection or transformation techniques. In this paper, band selection based technique is used. Inter band block correlation coefficient method followed by Singular Value Decomposition (SVD) and QR decomposition is utilized for reducing the dimension of a hyperspectral image.

2.1. Inter Band Block Correlation Coefficient

Inter Band Block Correlation Coefficient (IBBC)[4] is a preprocessing algorithm to remove the water absorption and noisy bands automatically. This technique reduces the computational complexity of band selection techniques.

Consider a hyperspectral image of dimension $m \times n \times l$, where l is the number of bands. Each band is divided into blocks of size $q \times q$. If m, n is not divisible by q, the bands are zero padded with $r = q - mod(m, q)$ rows and $c = q - mod(n, q)$ columns. So the size of bands become $M \times N$, where $M = m + r$ and $N = n + c$. Now, consider two consecutive bands and find the correlation between the block B_i of j^{th} band and $\widetilde{B_i}$ of $(j+1)^{th}$ band. The equation for the correlation coefficient is given by

$$\rho_j(i) = \frac{N_p(b_i^T \widetilde{b_i}) - (e^T b_i)(e^T \widetilde{b_i})}{\sqrt{[N_p(b_i^T b_i) - (e^T b_i)^2][N_p(\widetilde{b_i}^T \widetilde{b_i}) - (e^T \widetilde{b_i})^2]}}; \quad 1 \leqslant j \leqslant l-1,\ 1 \leqslant i \leqslant N_b \tag{1}$$

where b_i and $\widetilde{b_i}$ are the vector form or 1D version of the blocks B_i and $\widetilde{B_i}$ respectively, e is the column vector of ones of size $N_p \times 1$, $N_p = q^2$ is the number of pixels in each block, $N_b = \frac{M}{q} \times \frac{N}{q}$ and $\rho_j \in \mathbb{R}^{N_b}$. The correlation matrix for all the blocks in consecutive bands is given by $\rho = [\rho_1, \rho_2, \ldots, \rho_{(l-1)}]$. Standard deviation of the correlation matrix is found out and represented in vector form as $\sigma = [\sigma_1, \sigma_2, \ldots, \sigma_{l-1}]$ where,

$$\sigma_j = \sqrt{\frac{1}{N_b}(\rho_j - \mu_j e)^T(\rho_j - \mu_j e)}; \tag{2}$$

Here e is column vector of ones of size $N_b \times 1$ and $\mu_j = \frac{e^T \rho_j}{N_b}$ is the mean of the vector ρ_j.

A threshold is determined such that all the elements above this threshold are eliminated. Threshold can be based on mean, standard deviation and 2-mean. In this paper, standard deviation of σ is taken as the threshold value.

2.2. Singular Value Decomposition

Any matrix X of dimension $a \times b$ can be decomposed into matrices U, Σ and V^T. It is expressed as:

$$X = U\Sigma V^T \tag{3}$$

where U is an $a \times a$ orthogonal matrix with eigen vectors of XX^T in its columns, Σ is a $a \times b$ matrix with square root of eigen values of XX^T or X^TX as its diagonal elements and V is a $b \times b$ orthogonal matrix with eigen vectors of X^TX in its columns. In SVD based band selection, only the first k eigen values will be considered. ie, k most informative bands are selected and the rest are discarded.

2.3. QR Decomposition

QR decomposition (also called as QR factorization) is a process of orthogonalization which decomposes a matrix into an orthogonal matrix Q and an upper triangular matrix R. Consider a $a \times b$ matrix X with $a \geq b$, QR decomposition can be written as:

$$X = QR \tag{4}$$

where Q is a $a \times a$ matrix and R is a $a \times b$ matrix. Here the diagonal values of R are not arranged in order. To reorder this in descending order, a permutation matrix P is used such that:

$$XP = QRP = Q\widehat{R} \tag{5}$$

3. Kernel Based Classification

In this section, the two kernel based classification techniques used here, Support Vector Machine (SVM)[10] and Regularized Least Square(RLS) are briefly explained.

3.1. Regularized Least Square

The main goal of Regularized Least Square (RLS)[11] is to minimize the L2 norm of error and there by calculating the weight matrix. This weight matrix is further utilized for predicting the labels of the testing samples.

Consider a training set with training samples $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ and training labels $(y_1, y_2, \ldots, y_n)$ where $x_i \in \mathbb{R}^d$, $y_i \in \{1, 2, \ldots, T\}$ for $i = 1, 2, \ldots, n$. $\mathbf{Y}$ is an $n \times T$ output matrix with $Y_{ij} = 1$ if i^{th} training sample belongs to j^{th} class and -1 otherwise. The optimization problem for a linear model can be formulated as:

$$\min_{W \in \mathbb{R}^{d \times T}} \left\{ \frac{1}{n} \|\mathbf{Y} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \right\} \tag{6}$$

where $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}]^T$ is a $n \times d$ matrix. In the case of a non-linear model, the above formulation can be modified as:

$$\min_{C \in \mathbb{R}^{n \times T}} \left\{ \frac{1}{n} \|\mathbf{Y} - \mathbf{KC}\|_F^2 + \lambda \mathbf{C}^T \mathbf{KC} \right\}. \tag{7}$$

where $\mathbf{K}$ is an $n \times n$ matrix which includes the kernel functions $K_{ij} = k(\mathbf{x_i}, \mathbf{x_j})$. Optimum value of C can be calculated and is used for prediction of class label.

Grand Unified Regularized Least Squares (GURLS) is a software package used in various supervised learning problems. It utilizes RLS for its functioning and has a great importance in large scale multi class problems.

3.2. Support Vector Machine (SVM)

SVM[6] is a supervised machine learning technique used for classification and regression. It was originally designed to serve binary classification. Consider a binary class linearly separable data with training set (x_i, y_i) where $x_i \in \mathbb{R}^n$ is the training sample, $y_i \in (-1, 1)$ is the training label and $i = 1, 2, \ldots, m$. The problem formulation of SVM in matrix format for this case is:

$$\begin{gathered} \min_{w,\xi} \ \frac{1}{2} w^T w \\ subject\ to: \ D(Xw - \xi e) \geqslant e \end{gathered} \tag{8}$$

where D is a diagonal matrix of size $m \times m$ with class labels as the diagonal elements, X is the $m \times n$ data matrix, e is $m \times 1$ column vector of ones, $w = [w_1, w_2, \ldots w_n]^T$ and $\xi \in \mathbb{R}$. In case of non-linear data, a mapping function ϕ is utilized to map the data to higher dimension.

LibSVM is a kernel based software library which utilizes multiclass SVM[12] for classification and regression. Usually for a multiclass problem, many binary classifiers are constructed. This can be achieved by one-against-one or one-against-all techniques. In LibSVM, one-against-one method is used to generate $\frac{k(k-1)}{2}$ binary classes for a k-class problem.

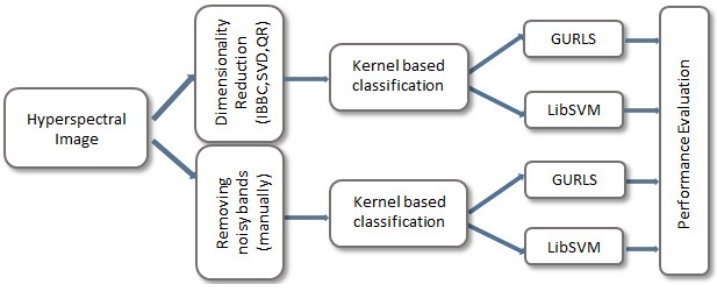

Fig. 1: Flow diagram of important processes in the proposed work.

4. Methodology

The experiment is performed on standard Indian Pines and Salinas-A datasets. The image is captured using the 224 band sensor AVIRIS which operates in visible to infrared region in the electromagnetic spectrum. Indian Pines dataset consists of 16 classes spread across 145×145 pixels of each band. Salinas-A is a subset of Salinas image (within the region (591-676, 158-240)) which comprises of 6 classes spread across 86×83 pixels. The common experimental procedure is to manually remove the noisy and water absorption bands and using the remaining bands for classification. In this paper, Inter band block correlation coefficient method is used to automatically select and remove the noisy bands. A significant improvement in the computational complexity is achieved on applying the proposed method to the dataset before performing its singular value decomposition. The preprocessed data is then given for dimensionality reduction using SVD. The output matrix V of SVD is given as an input to the QR decomposition algorithm. Kernel based classification is done using GURLS and LibSVM libraries on the original and dimensionally reduced data. The hyperspectral cube of size $m \times n \times l$ is converted into a two dimensional data of size $l \times mn$. 10, 20, 30, 40,and 50% of pixel vectors from each class are chosen for training and all the pixel vectors in the data are taken for testing.Before generating training and testing samples, the background pixels (class 0) are removed. Accuracy assessment measures like Class wise accuracy (CA), Overall accuracy (OA), Kappa coefficients (K) and Average accuracy (AA) are used to validate the performance of classification with and without dimensionality reduction. These measures are calculated from the confusion matrix which is generated using original class label and predicted class label. The flow graph of these steps is shown in fig.1.

5. Experiments and Result Analysis

The water absorption bands are removed manually from Indian Pines and Salinas-A datasets. This is achieved by discarding the bands [104-108], [150-163], 220 of Indian pines and bands [108-112], [154-167], 224 of Salinas-A dataset. The remaining bands are used for classification in GURLS and LibSVM packages. In GURLS, classification was done using LOO (leave one out) and HO (hold out) cross validations in linear, radial basis function (RBF) and randfeats kernels. In this library, the parameters are selected automatically. The classification in LibSVM was performed using linear, polynomial and RBF kernels with 5-fold cross validation. In this paper, the control parameters C and γ are chosen by trial and error method such that these parameters give the best cross validation accuracy. C and γ parameters were estimated to be 1000, 0.4 for Indian pines and 100, 0.5 for Salinas-A dataset.

Inter band block correlation coefficient is performed on the hyperspectral image and the remaining 188 bands of Indian Pines and 183 bands of Salinas-A dataset are given for SVD and QR decomposition. After dimensionality reduction, only 100 bands are given for classification. C and γ parameters for dimensionally reduced datasets were estimated to be 10^6, 0.6 for Indian pines and 100,0.5 for Salinas-A. For 10% training sample, the overall accuracy for original and dimensionally reduced Indian pines dataset are 83.52% and 84.67% respectively. Similarly, for Salinas-A dataset the overall accuracies are 99.10% and 99.3082%. From this experiment it is clear that even with 50% reduction in dimension, comparable accuracies are obtained by kernel based classification. Further to validate the analysis, the experiment was extended to 20, 30, 40 and 50% training samples. The classification accuracies for 10 and 40 % training samples of original and dimension reduced datasets are tabulated in Table1 and Table 2,3 respectively. Figure

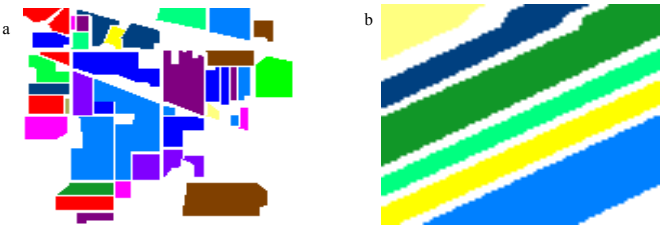

Fig. 2: Ground Truth (a) Indian Pines (b) Salinas-A

Table 1: Classification Accuracies of 10% and 40% Indian Pines and Salinas-A dataset for GURLS and LibSVM

% of training data	Accuracies	GURLS (RBF kernel)		LibSVM (Polynomial kernel)	
		Indian Pines	Salinas-A	Indian Pines	Salinas-A
10	OA	83.52	99.5325	80.67	99.1025
	AA	77.18	99.40	65.31	98.96
	K	0.811	0.9941	0.777	0.9888
40	OA	93.65	99.8878	91.39	99.9065
	AA	92.53	99.88	88.26	99.89
	K	0.928	0.9986	0.901	0.9988

Table 2: Classification Accuracies of dimensionally reduced IndianPines dataset for GURLS and LibSVM

% of training data	Accuracies	GURLS						LibSVM		
		Linear		RBF		Randfeat		Linear	RBF	Polynomial
		CV=loo	*CV=ho*	*CV=loo*	*CV=ho*	*CV=loo*	*CV=ho*	*CV=5-fold*	*CV=5-fold*	*CV=5-fold*
10	OA	66.8065	66.8065	84.0472	**84.6717**	49.2341	69.7044	79.4321	80.3103	**82.8569**
	AA	50.83	50.83	81.10	**82.28**	28.49	53.04	78.84	79.66	**80.47**
	K	0.6098	0.6098	0.8176	**0.8247**	0.3768	0.6452	0.7652	0.7757	**0.8038**
40	OA	68.5335	68.5433	93.3554	**93.5799**	51.1465	69.9953	87.2280	**89.5697**	88.6818
	AA	51.41	51.42	92.64	**92.81**	31.28	52.33	88.50	**90.76**	90.24
	K	0.6286	0.6287	0.9241	**0.9267**	0.4034	0.6431	0.8542	**0.8811**	0.8707

2 shows the ground truth for Indian Pines and Salinas-A datasets. Classification map for 10% training samples of Salinas-A and 10, 30, 50% training samples of Indian pines dataset with dimensional reduction are as shown in fig.5 and fig.4 respectively. Comparison of the Indian pines and Salinas-A datasets with and without reducing dimension is shown in fig.3.

6. Conclusion

In this paper, the dimension of Indian Pines and Salinas-A datasets are reduced effectively using band selection techniques. Dimensionality reduction reduces the computational complexity while preserving the critical information in the hyperspectral images. Experimental analysis shows that even with 50% reduction in dimension, the accuracies obtained from kernel based classification in dimensionally reduced image is comparable to that of the original image.

Table 3: Classification Accuracies of dimensionally reduced Salinas-A dataset for GURLS and LibSVM

% of training data	Accuracies	GURLS						LibSVM		
		Linear		RBF		Randfeat		Linear	RBF	Polynomial
		CV=loo	*CV=ho*	*CV=loo*	*CV=ho*	*CV=loo*	*CV=ho*	*CV=5-fold*	*CV=5-fold*	*CV=5-fold*
10	OA	94.6522	94.3717	**99.3269**	99.009	98.3358	98.7472	99.2895	39.7158	**99.3082**
	AA	93.03	92.63	**99.18**	98.81	97.82	98.36	99.16	29.95	**99.16**
	K	0.9326	0.9290	**0.9916**	0.9876	0.9791	0.9843	0.9911	0.1721	**0.9913**
40	OA	92.9693	95.6619	**99.9252**	99.8504	98.6911	98.9529	99.9626	68.4929	**99.9813**
	AA	90.70	94.36	**99.92**	99.85	98.34	98.66	99.94	63.91	**99.96**
	K	0.9111	0.9454	**0.9991**	0.9981	0.9836	0.9869	0.9995	0.5866	**0.9998**

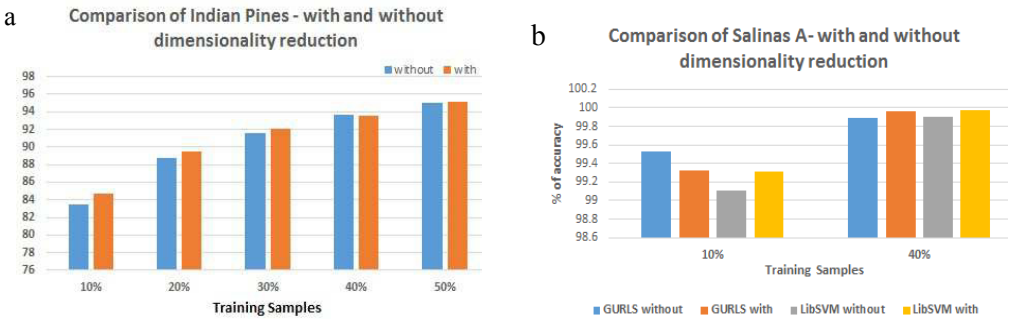

Fig. 3: Comparison of Overall accuracy of the datasets with and without dimensionality reduction (a) Indian Pines (b) Salinas-A

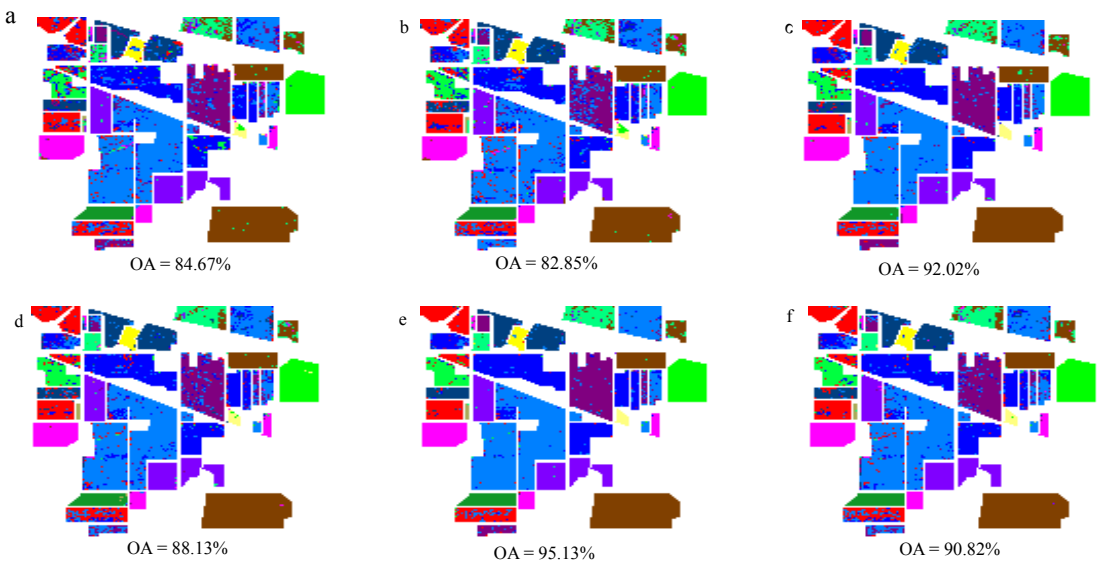

Fig. 4: Classification maps of dimensionally reduced Indian Pines data set for 10, 30, 50% training data with classifiers: GURLS - RBF kernel and LibSVM - polynomial kernel (a) GURLS-10% (b) LibSVM-10% (c) GURLS-30% (d) LibSVM-30% (e) GURLS-50% (f) LibSVM-50%

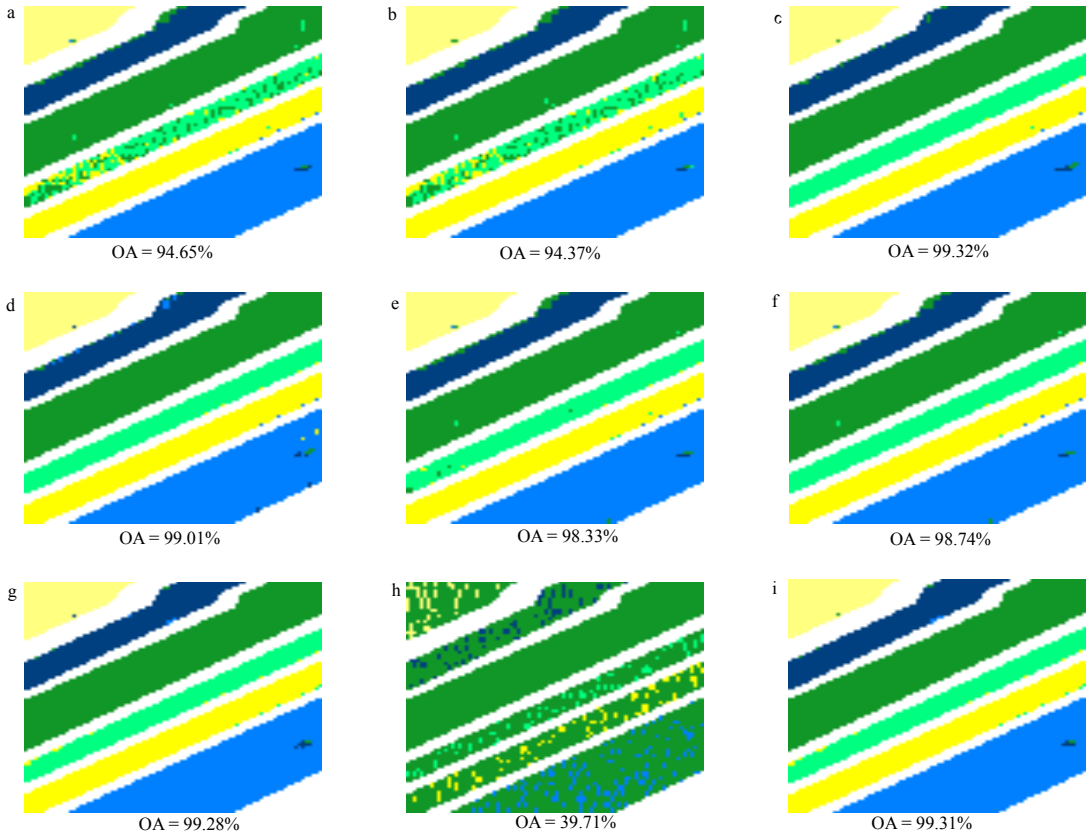

Fig. 5: Classification maps of dimensionally reduced Salinas-A data set for 10% training data with classifiers: GURLS a) linear-loo b) linear-ho c) rbf-loo d) rbf-ho e) randfeat-loo f) randfeat-ho and LibSVM g) linear h) rbf i) polynomial

References

1. Harsanyi, J.C., Chang, C.I.. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *Geoscience and Remote Sensing, IEEE Transactions on* 1994;**32**(4):779–785.
2. Bruce, L.M., Koger, C.H., Li, J.. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *Geoscience and Remote Sensing, IEEE Transactions on* 2002;**40**(10):2331–2338.
3. Lodha, S.P., Kamlapur, S.. Dimensionality reduction techniques for hyperspectral images. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)* 2014;**3**(10).
4. Bhushan, D.B., Sowmya, V., Manikandan, M.S., Soman, K.. An effective pre-processing algorithm for detecting noisy spectral bands in hyperspectral imagery. In: *Ocean Electronics (SYMPOL), 2011 International Symposium on*. IEEE; 2011, p. 34–39.
5. Schölkopf, B., Smola, A.J.. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press; 2002.
6. Soman, K., Loganathan, R., Ajay, V.. *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd.; 2009.
7. Haridas, N., Sowmya, V., Soman, K.. Gurls vs libsvm: Performance comparison of kernel methods for hyperspectral image classification. *Indian Journal of Science and Technology* 2015;**8**(24).
8. Tacchetti, A., Mallapragada, P.K., Santoro, M., Rosasco, L.. Gurls: a least squares library for supervised learning. *The Journal of Machine Learning Research* 2013;**14**(1):3201–3205.
9. Chang, C.C., Lin, C.J.. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2011;**2**(3):27.
10. Gualtieri, J., Chettri, S.R., Cromp, R., Johnson, L.. Support vector machine classifiers as applied to aviris data. In: *Proc. Eighth JPL Airborne Geoscience Workshop*. 1999, .
11. Tacchetti, A., Mallapragada, P.S., Santoro, M., Rosasco, L.. Gurls: a toolbox for regularized least squares learning 2012;.
12. Hsu, C.W., Lin, C.J.. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on* 2002; **13**(2):415–425.