Indian Journal of Science and Technology, *Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/99594, September 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Improvement in Kernel based Hyperspectral Image Classification Using Legendre Fenchel Denoising

R. Reshma*, V. Sowmya and K. P. Soman

Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore- 641112, Tamil Nadu, India; reshmaraju2992@gmail.com, v_sowmya@cb.amrita.edu, kp_soman@amrita.edu

Abstract

Hyperspectral images have bulk of information which are widely used in the field of remote sensing. One of the main problems faced by these images is noise. This emphasizes the importance of denoising techniques for enhancing the image quality. In this paper, Legendre Fenchel Transformation (LFT) is used for preprocessing the Indian Pines Dataset. LFT reduces the noise of each band of the hyperspectral image without affecting the edge information. Signal to noise ratio is computed which helps to evaluate the performance of denoising. Further, the denoised image is classified using GURLS and LibSVM and the various accuracies are estimated. The experimental analysis shows that the overall and classwise accuracies are more for the preprocessed data classification when compared to the classification without preprocessing. The classification accuracy is improved with denoising of hyperspectral image.

Keywords: Classification, Denoising, GURLS, Hyperspectral Image, Kernel Methods, Legendre Fenchel, LibSVM

1. Introduction

Hyperspectral imaging involves acquiring images of earth's surface in innumerable spectral bands. With the abundant information available in hundreds or thousands of bands, which range from visible to infrared region, the hyperspectral images (HSI) have increasing applications in fields like environmental monitoring. Some of the main challenges faced by HSI are noises, higher dimensional data etc. This reveals the importance of applying preprocessing techniques in these images.

Hyperspectral denoising is one of the preprocessing approaches to reduce the noises that affects the quality of the captured image. Various denoising techniques include Principal Component Analysis (PCA), Total variation, Least square, Wavelet, Singular Valued Decomposition, Legendre Fenchel Transformation[1,2] etc. The main advantage of LFT is that it preserves the edge information of the hyperspectral images. This paper utilizes LFT for denoising which converts the original function in the primal form to the dual form which enhances the computation time and the ease of divergence.

Hyperspectral classification techniques are applied to the preprocessed HSI. Orthogonal Matching Pursuit (OMP), Regularized Least Square (RLS), Support Vector Machine (SVM)[3], Subspace Pursuit (SP), Relevance Vector Machine (RVM), etc. are the commonly used methods for classification. In this paper, classification is performed using kernel based libraries like Grand Unified Regularized Least Squares (GURLS) [4] and LIBSVM [5,6].

This paper mainly deals with the improvement in the HSI quality by Legendre Fenchel denoising technique. Signal to noise ratio (SNR) is used to show the enhancement in the image after denoising. Comparison of the classification accuracies of the HSI data with and without denoising is discussed.

2. Legendre Fenchel denoising

Legendre Fenchel Transformation (LFT) maps the primal form $(x, h(x))$ to the dual form $(a, h^*(a))$. This technique utilizes the duality property to upgrade the divergence rate and computational time. Consider a continuous function $h : R \rightarrow R$, then LFT can be expressed as:

**Author for correspondence*

$$h^*(a) = \sup_{x \in R} \{ax - h(x)\} \tag{1}$$

Here a is the slope of line passing through the point $(x, h(x))$ which has maximum intercept on y-axis. A dual ROF model can be formulated based on this concept. The standard expression for a ROF model[7] is given by:

$$\min_v \|\nabla v\|_1 + \frac{\lambda}{2}\|v - g\|_2^2 \tag{2}$$

where, v is the denoised image, g is the observed image and λ is the regularization parameter. Using the LF transform, the dual form of ROF model can be formulated as:

$$\min_v \max_{a := \|a\|_\infty \le 1} \left(\langle a, \nabla v \rangle - \delta_a(\mathrm{A}) + \frac{\lambda}{2}\|v - g\|_2^2 \right) \tag{3}$$

where,

$$\|\nabla v\|_1 = \max_{a \in A} \left(\langle a, \nabla v \rangle - \delta_a(A) \right) \tag{4}$$

$$\delta(a) = \begin{Bmatrix} 0 & if\ \|a\| \le 1 \\ \infty & otherwise \end{Bmatrix} \tag{5}$$

Here, A is the indicator function and we have to find the update equations for variables v and a. The update equations are given by:

$$a^{m+1} = \frac{a^m + \sigma \nabla v^m}{\max(1, |a^m + \sigma \nabla v^m|)} \tag{6}$$

$$v^{m+1} = \frac{v^m + \tau\, diva^{m+1} + \tau\lambda g}{1 + \tau\lambda} \tag{7}$$

where, τ is Lipchitz constant, m is the number of iteration and λ is control parameter. Computational complexity of the process is reduced using nabla matrix.

3. Kernel based Classification

The two kernel based[8] classification techniques used in this paper are Support Vector Machine (SVM) and Regularized Least Square (RLS).

3.1 GURLS

Grand Unified Regularized Least Squares is a software library which utilizes regularized least square (RLS) technique for multi class classification and regression. Various kernels used in this include linear, randfeats and radial basis function (RBF). RLS minimizes the L2 norm of error and calculates the weight matrix.

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

$$\min_{C \in \mathbb{R}^{n \times T}} \left\{ \frac{1}{n}\|\mathbf{Y} - \mathbf{KC}\|_F^2 + \lambda \mathbf{C}^T\mathbf{KC} \right\}. \tag{8}$$

Optimum value of C can be calculated. If a linear model is considered, then the formulation can be modified as:

$$\min_{W \in \mathbb{R}^{d \times T}} \left\{ \frac{1}{n}\|\mathbf{Y} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \right\}, \tag{9}$$

where, $\mathbf{X} = [x_1, x_2, \ldots, x_n]^T$ is a $n \times d$ matrix. Using this optimum value, the class labels for the testing samples are estimated.

3.2 LibSVM

LibSVM is a kernel based library for Support Vector Machines which is used for many machine learning tasks. SVM[9] is a supervised machine learning[10] algorithm for regression and classification. The original design of SVMs serves the binary classification. So for multi-class classification, many binary classifiers are generated. This can be done in two ways: one-against-all and one against one. In LibSVM, one-against-one method of multiclass SVM is used. For this method, $k(k-1)/2$ binary classes are generated from a k-class problem. The kernels in LibSVM include linear, polynomial, RBF and sigmoid.

Consider a linearly separable binary class data with training set (x_i, y_i) where $x_i \in R^n$ is the training sample, $y_i \in (-1,1)$ is the training label and $i=1,2,\ldots m$. The hyperplane that separates the two classes is given by:

$$w^T x_i - \xi = 0 \quad , \tag{10}$$

where, $w = [w_1, w_2, \ldots, w_n]^T$ and $\xi \in R$. So the decision function will be $sign(w^T x_i - \xi)$. The problem formulation of SVM in matrix format for this case is:

$$\min_{w,\xi} \frac{1}{2} w^T w \quad , \tag{11}$$

$$subject\ to : D(Xw - \xi e) \ge e$$

where X is data matrix of size $m \times n$, D is $m \times m$ diagonal matrix with class labels as the diagonal elements and e is a column vector of ones with size $m \times 1$. For a non-linear data, the data will be first mapped to a higher dimension using function ϕ.

4. Method ology

The experiment is performed on Indian Pines hyperspectral dataset. The image is captured using the sensor AVIRIS which operates in visible to infrared region in the electromagnetic spectrum. The dataset consists of 16 classes spread across 145×145 pixels of each band. The information is spread across 224 bands in the original Indian pines dataset. The common experimental procedure is to manually remove the 24 water absorption bands and using the remaining bands for classification. In this paper, Legendre Fenchel denoising technique is done and then the preprocessed data is given for kernel based classification techniques like RLS (in GURLS) and SVM (in LibSVM) as shown in Figure 1.

Preprocessing technique, namely Legendre Fenchel denoising is applied to the original Indian Pines dataset. The performance of this method is first visually evaluated and then it is assessed by comparing the signal to noise ratio of the denoised image and the original image. SNR[11] for each band of a HSI is calculated using the expression:

$$SNR_b = 10\log_{10}\left(\frac{\sum_{ij} a_{ijb}^2}{\sum_{ij}\left(a_{ijb}-\mu_b\right)^2}\right) \tag{12}$$

where, a_{ijb} is the pixel value at (i,j) in band b and μ_b is the mean of a_{ijb} in a homogeneous area. Such an area contains all pixels which belongs to the same class. The SNR value varies with different homogeneous area. So to overcome this, SNR is calculated as the mean of the SNR of homogeneous areas of different classes.

The hyperspectral cube of size $m \times n \times b$ is converted into a two dimensional data of size $b \times mn$ where b denotes the number of bands. Each column of this 2D data represents the pixel vectors. 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 are removed.

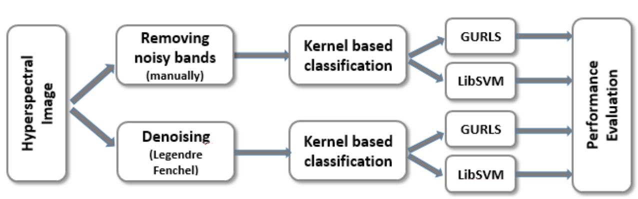


Figure 1. Flow diagram of important processes in the proposed work.

Accuracy assessment measures[12] are used to validate the performance of classification with and without preprocessing. Class wise accuracy (CA), Overall accuracy (OA), Kappa coefficients (K) and Average accuracy (AA) are calculated from the confusion matrix generated. Confusion matrix is formed using original class label and predicted class label.

5. Experiments and Discussions

The proposed method is performed in Indian Pines dataset. The Legendre Fenchel Transformation is done to

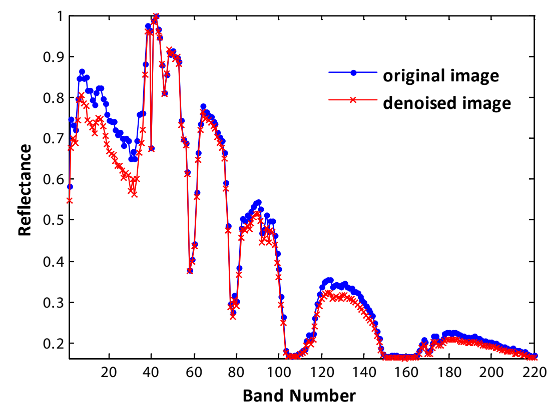


Figure 2. Spectral reflectance of pixel (10,10) across the 220 bands of the Indian pines dataset.

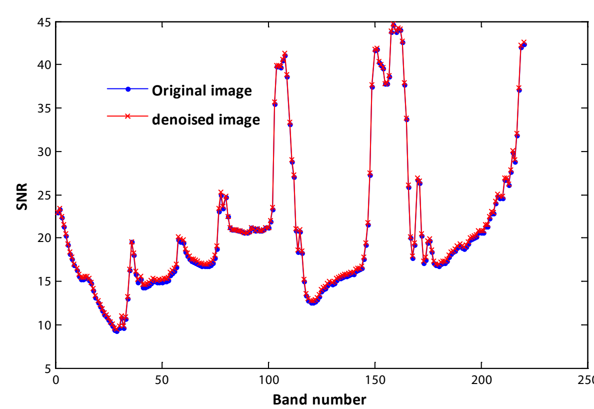


Figure 3. SNR across the 220 bands of the Indian pines dataset.

Table 1. Classification accuracies of Gurls and Libsvm for 10% training dataset of denoised Indian pines subset

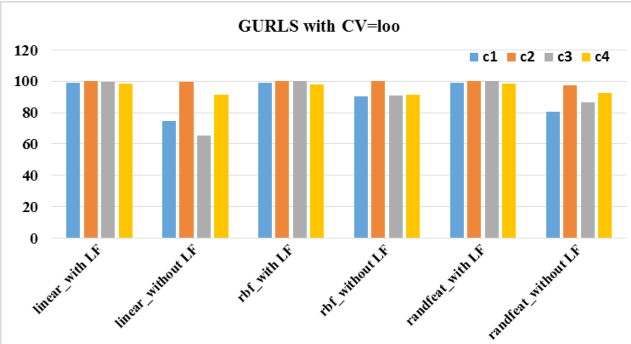
Class Label	Class Name	*GURLS*						*LibSVM*		
		Linear		*RBF*		*Randfeat*		*Linear*	*RBF*	*Poly*
		CV=LOO	*CV=HO*	*CV=LOO*	*CV=HO*	*CV=LOO*	*CV=HO*	*CV=5-fold*	*CV=5-fold*	*CV=5-fold*
C1	Corn-notill	99.20	99.20	99.20	99.20	99.00	99.40	98.71	98.81	98.81
C2	Grass-trees	100	100	100	99.86	100	99.86	100	100	100
C3	Soybean-notill	99.73	99.73	100	99.73	100	99.73	100	100	99.86
C4	Soybean-mintill	98.74	98.63	98.21	98.84	98.42	98.42	98.27	98.48	98.53
Overall Accuracy		99.2220	99.1762	99.0389	**99.2449**	99.0847	99.1076	98.9474	99.0618	99.0618
Average Accuracy		99.4175	99.39	99.3525	99.4075	99.355	99.3525	99.245	99.3225	99.30
Kappa Coefficient		0.9889	0.9883	0.9863	**0.9892**	0.9847	0.9873	0.9850	0.9866	0.9866

Table 2. Overall accuracy, average accuracy and kappa coefficient of Gurls and Libsvm for 10, 20, 30, 40, 50% training dataset of denoised Indian pines subset

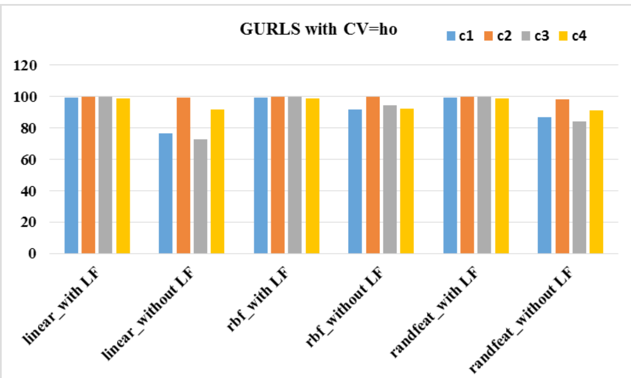
% of training data	Accuracies	*GURLS*						*LibSVM*		
		Linear		*RBF*		*Randfeat*		*Linear*	*RBF*	*Poly*
		CV=LOO	*CV=HO*	*CV=LOO*	*CV=HO*	*CV=LOO*	*CV=HO*	*CV=5-fold*	*CV=5-fold*	*CV=5-fold*
10	OA	99.2220	99.1762	99.0389	**99.2449**	99.0847	99.1076	98.9474	99.0618	99.0618
	AA	**99.4175**	99.39	99.3525	99.4075	99.355	99.3525	99.245	99.3225	99.30
	K	0.9889	0.9883	0.9863	**0.9892**	0.9847	0.9873	0.9850	0.9866	0.9866
20	OA	99.4050	99.3135	99.6110	99.4508	99.1076	99.4279	99.6796	**99.611**	99.6568
	AA	99.5875	99.5225	99.67	99.555	99.37	99.6	99.76	**99.71**	99.73
	K	0.9915	0.9902	0.9945	0.9922	0.9873	0.9919	0.9954	**0.9945**	0.9951
30	OA	99.4508	99.4050	**99.7712**	99.5652	99.2220	99.4737	99.5881	99.6568	99.5881
	AA	99.5325	99.495	**99.7975**	99.6325	99.365	99.5125	99.66	99.73	99.66
	K	0.9922	0.9915	**0.9967**	0.9938	0.9889	0.9925	0.9941	0.9951	0.9941
40	OA	99.5195	99.5423	**99.8627**	**99.8627**	99.3822	99.5423	99.6110	99.6568	99.5423
	AA	99.605	99.6175	**99.8725**	**99.8725**	99.47	99.61	99.69	99.72	99.61
	K	0.9932	0.9935	**0.9980**	**0.9980**	0.9912	0.9935	0.9945	0.9951	0.9935
50	OA	99.6339	99.5423	**99.9771**	99.8169	99.5881	99.6568	99.7941	99.8169	99.7941
	AA	99.7075	99.63	**99.975**	99.8	99.645	99.685	99.80	99.84	99.81
	K	0.9948	0.9935	**0.9997**	0.9974	0.9941	0.9951	0.9971	0.9974	0.9971

all bands of the hyperspectral image. Different parameters of this algorithm like number of iterations n, control parameter λ, and Lipchitz constant τ are determined by trial and error method and is visually evaluated to be $n=40$, $\lambda=15$, $\tau=\sqrt{8}$.

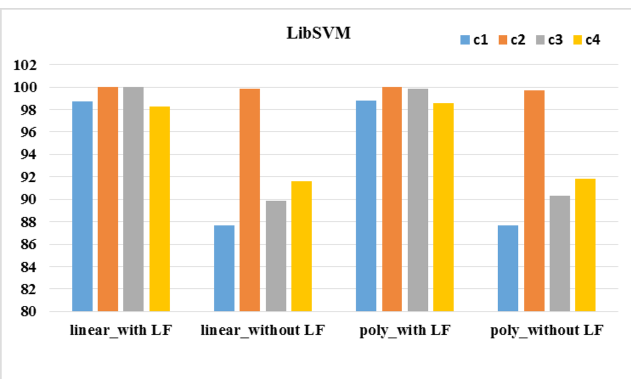
The changes in reflectance value of (10,10) pixel in original and denoised images can be seen in the reflectance v/s band number graph shown in Figure 2. The denoised data is given for classification. Figure 6 shows the image of band 3 in Indian Pines dataset before and after denoising. SNR value of the different bands before and after denoising are calculated using eq.12. The region highlighted in Figure 6a) is an example of homogeneous area. Figure 3 shows the plot of SNR value of original and denoised image v/s band number. From this plot it is clear that the SNR value gets enhanced



(a)



(b)



(c)

Figure 4. Comparison of the classwise accuracies (using (a) GURLS with CV=loo (b) GURLS with CV=ho (c) LibSVM with 5-fold CV for classification) of the Indian Pines dataset with and without denoising.

after denoising which results in the improved quality of image.

The subset (31:116, 27:94) of the denoised image is given for classification. Figure 8 shows the ground truth for Indian Pines subset. Here C1, C2, C3 and C4 represent the four classes as mentioned in TABLE I. In the experiments we conducted, GURLS and LibSVM were the kernel based libraries used for classification. In GURLS both LOO (Leave One Out) and HO (Hold

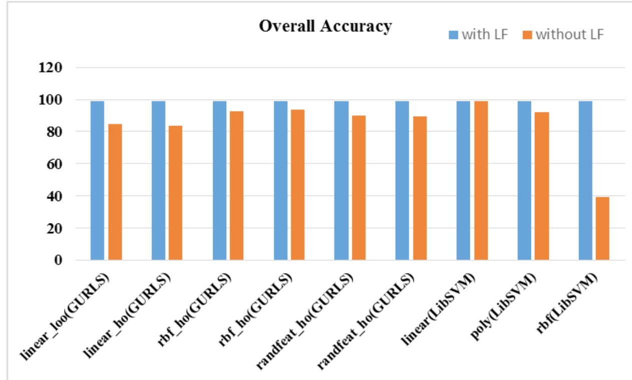


Figure 5. Comparison of the overall accuracies of the Indian Pines dataset with and without denoising.

(a)

(b)

Figure 6. Denoising result on band 3 of Indian Pines dataset. a) Original Image and the highlighted portion is a homogeneous area to find SNR b) Denoised image.

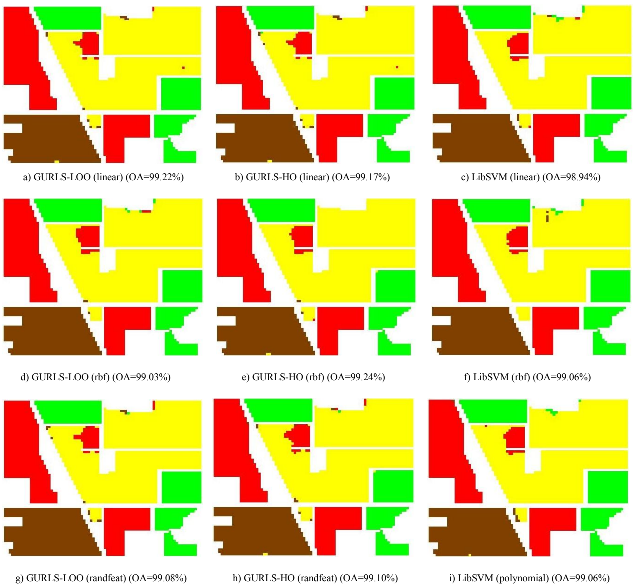


Figure 7. Classification map for 10% training samples of denoised Indian Pines dataset.

Figure 8. Ground truth for Indian Pines subset.

Out) cross validations were performed on the linear, RBF and randfeats kernels. Optimal parameters are automatically selected in this library. LibSVM utilizes k-fold cross validation. Here the data is divided into k subsets of almost equal size and one such subset is used for testing while the rest is used as training data. The process continues for k times so that all the subsets will be used for testing at least once. The control parameters C and γ are estimated to be 10^6 and 0.9 respectively by performing 5-fold cross validation. Classification in LibSVM was done using linear, RBF and polynomial kernels. Experiments were conducted on 10,20,30,40 and 50% of training data. Various accuracy assessment measures are utilized to estimate the performance of classification. Output of different accuracy assessment measures for classification of 10% training dataset is tabulated in TABLE I. Overall accuracy, Average accuracy and Kappa coefficient of the subset image of denoised Indian Pines dataset for 10%, 20%, 30%, 40% and 50% of training data are shown in TABLE II where the highlighted values are the best accuracy in each case. The classification map for 10% dataset is shown in Figure 7.

Classification results of 10% training data and the corresponding accuracy values of the Indian pines dataset without preprocessing or denoising is obtained[13]. Comparison of class wise accuracies of 10% training dataset with and without denoising is shown in Figure 4. Similarly, the overall accuracy comparison is graphically plotted in Figure 5. It is clear that there is a drastic increase in the accuracies of all classes after denoising. Classification accuracy of each class for a denoised dataset ranges from 98-100%. The improvement of OA is from 93.68% to 99.2449%, AA is from 94.42% to 99.41% and Kappa is from 0.9101 to 0.9892 for 10% training data. We know that classification accuracies increase with increase in percentage of training data. So the improvement in the accuracy and image quality with the Legendre Fenchel denoising technique is applicable for 20, 30, 40 and 50% of training samples as well.

6. Conclusion

In this paper, the effectiveness of applying denoising technique, namely Legendre Fenchel Transformation on the standard hyperspectral image, Indian pines is studied. The success of denoising was first examined using the signal to noise ratio of the image. SNR value increases with decrease in noise content. There was a notable increase in the classification accuracy of the denoised image. Experimental analysis shows the potency of applying denoising technique rather than removing the noisy bands manually.

7. References

1. Santhosh S, Abinaya N, Rashmi G, Sowmya V, Soman KP. A novel approach for denoising coloured remote sensing image using Legendre Fenchel Transformation. 2014 International Conference on Recent Trends in Information Technology (ICRTIT); 2014 Apr. p. 1–6.
2. Handa A. Applications of Legendre-Fenchel transformation to computer vision problems. Department of Computing at Imperial College London; 2011 Jul. p. 1–35.
3. Soman KP, Loganathan R, Ajay V. Machine learning with SVM and other Kernel methods. PHI Learning Pvt. Ltd; 2009 Feb.

4. Tacchetti A, Mallapragada PK, Santoro M, Rosasco L. Gurls: a least squares library for supervised learning. The Journal of Machine Learning Research. 2013 Oct; 14(1):3201–5.
5. Chang CC, Lin CJ. Libsvm: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011 Apr; 2(3).
6. Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks. 2002 Mar; 13(2):415–25.
7. Rudin LI, Stanley O, Emad F. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena. 1992 Nov; 60(1–4):259–68.
8. Scholkopf B, Smola AJ. Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press, Cambridge; 2001.
9. Gualtieri J, Chettri SR, Cromp R, Johnson L. Support vector machine classifiers as applied to aviris data. Proceeding Eighth JPL Airborne Geoscience Workshop, Citeseer; 1999 Feb.
10. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques; 2007 Jun. p. 249–68.
11. Xu L, Fan L, Wong A, Clausi DA. Hyperspectral image denoising using a spatial–spectral monte carlo sampling approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015 Jun; 8(6):3025–38.
12. Aswathy C, Sowmya V, Soman KP. ADMM based hyperspectral image classification improved by denoising using Legendre Fenchel transformation. Indian Journal of Science and Technology. 2015 Sep; 8(24):1–9. DOI: 10.17485/ijst/2015/v8i24/80030.
13. Haridas N, Sowmya V, Soman KP. Gurls vs Libsvm: Performance comparison of kernel methods for hyperspectral image classification. Indian Journal of Science and Technology. 2015 Sep; 8(24):1–10. DOI: 10.17485/ijst/2015/v8i24/80843.