A Comparative Analysis of Classification Techniques using Hyperspectral Data

Raghvendra Singh[1], Dr. Sunil Kumar Nimesh[2] and Pawan Kumar Tiwari[3]

1. *Junior Cartographic Assistant, Soil and Land Use Survey of India, Ministry of Agriculture & Farmers Welfare, Government of India, C - 4, Sector - 1, Noida (U.P.) – 201301 India,*
2. *Associate Professor, Department of Geography, National Post Graduate College, Bhongaon, Mainpuri – (U.P.) – 205262 India,*
3. *GIS Expert, SAI Consulting Engineers Pvt. Ltd. SAI House, Satyam Corporate Square, B/h. Rajpath Club, Bodakdev, Ahmedabad – 380059, India,*

ABSTRACT

The aim of study is to apply different techniques to the classification using hyperspectral images. The Hyperion instrument provides a new class of Earth observation data for improved Earth surface characterization. The Hyperion provides a high resolution hyperspectral imager capable of resolving 242 spectral bands (from 0.4 to 2.5 µm) with a 30-meter resolution. The instrument can capture (7.5 X 100) km land area per image and provide detailed spectral mapping across all 242 channels with high radiometric accuracy. The study is mainly focused on investigating the existing techniques of classification, both supervised and unsupervised of remotely sensed hyperspectral data sets with focus on Spectral Angle Mapper (SAM), Support Vector Machine (SVM) and active learning techniques, evaluating their advantages and disadvantages and to provide a set of recommendations for the new classification techniques developed, after carefully evaluating and assessing their performance in terms of classification accuracy.

Furthermore NDVI image of the study area is also generated using Hyperion hyperspectral bands and results were compared among all classification method used. Classification techniques used in the study succeeded in taking advantage of the spatial and the spectral information simultaneously. All classification methods perform well with hyperspectral data in general and SVM classification technique proved more efficient compared to other methods with overall classification accuracy of 86.85%.

Key words: *Hyperspectral, Hyperion, Supervised/Unsupervised classification, SAM, SVM, NDVI.*

1. Introduction

The aim of work is to apply different techniques for the classification of hyperspectral images. A typical hyperspectral image has 30 to over 200 bands of data usually covering visual and other regions of the electromagnetic spectrum. The bands of data at each pixel form a spectrum that contains spectral features, which are significant relative differences in

neighboring spectral bands values. When the pure spectral signatures are identified, the proportion of each material in each pixel can be estimated. Abundances provide additional information about the composition of each pixel; if this information is used in a correct way, it may complement the results provided by traditional "hard" classification techniques.

2. Study area

The study was conducted in Delhi, comprising National Capital Region that lies between 28°35′11″ to 28°39′35″N, Longitude and 77°11′88″ to 77°16′41″ E Latitude. The annual temperature varies from 3°C to 45°C and annual rainfall ranges between 400 mm to 600 mm. Delhi is a densely populated state with mostly urban structures. The population of Delhi is 16.75 million (Census 2011). The forest cover in the state based on interpretation of satellite data (2008) is 176.20 km^2, which is 11.88% of the state's geographical area. In terms of forest canopy density classes, Delhi has 6.76 km^2 area under very dense forest, 49.48 km^2 area under moderately dense forest and 119.96 km^2 area under open forest (India State of Forest Report- Delhi, 2011).

3. Type of Classification

3.1 Supervised Classification

3.1.1 Spectral Angle Mapper (SAM) classification

The input file for classification must be an atmospheric corrected data. In this study QUAC corrected file was taken for the further analysis as it gives more corrected spectra rather than FLAASH. All the further classification steps were done with the QUAC corrected file.

Input requirement for SAM classification

Spectral Angle Mapper (SAM) is a physically-based spectral classification that uses an *n*-D angle to match pixels to reference spectra. The algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra and treating them as vectors in a space with dimensionality equal to the number of bands. It calculates the angular distance between each spectrum in the image and the reference spectra or "end members" in *n*-dimensions. The result is a classification image showing the best SAM match at each pixel and a "rule" image for each end-member showing the actual angular distance in radians between each spectrum in the image and the reference spectrum. Darker pixels in the rule images represent smaller spectral angles, and thus spectra that are more similar to the reference spectrum. The rule images can be used for subsequent classifications using different thresholds to decide which pixels are included in the SAM classification image.

Application of SAM algorithm results in the so called rule images, their values indicating spectral angles. On the other hand, linear unmixing results in abundance images. For final classification results, threshold value of 0.09 radian or less was used for SAM whereas abundance of 0.50 or more was selected for linear unmixing.

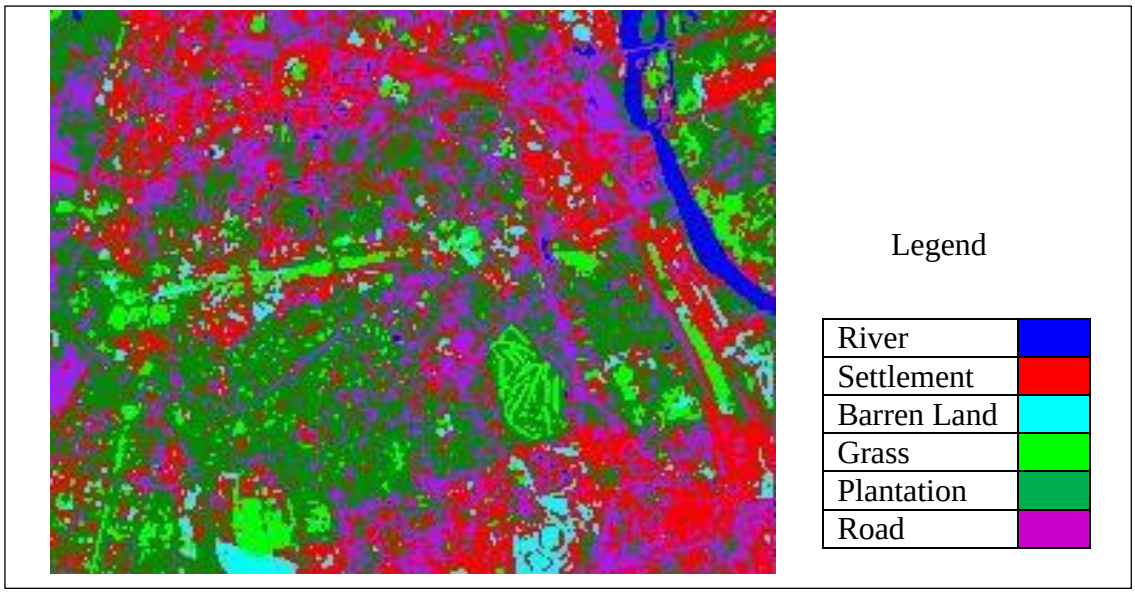

Figure 1: Spectral Angle Mapper (SAM) classification Map

3.1.2 Support Vector machine (SVM) Classification

Support Vector Machine (SVM) is a supervised classification method derived from statistical learning theory that often yields good classification results from complex and noisy data. It separates the classes with a decision surface that maximizes the margin between the classes. The surface is often called the optimal hyperplane, and the data points closest to the hyperplane are called support vectors. The support vectors are the critical elements of the training set.

SVM includes a penalty parameter that allows a certain degree of misclassification, which is particularly important for non-separable training sets. The penalty parameter controls the trade-off between allowing training errors and forcing rigid margins. It creates a soft margin that permits some misclassifications, such as it allows some training points on the wrong side of the hyperplane. Increasing the value of the penalty parameter increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well.

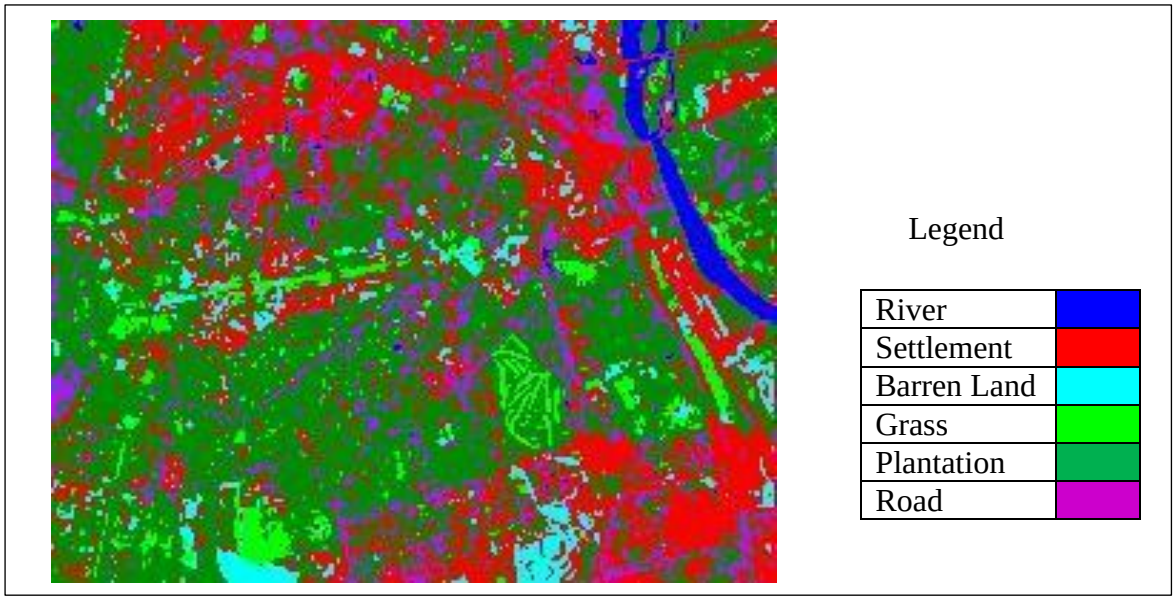

Figure 2: Support Vector machine (SVM) classification Map

3.1.3 Artificial Neural Network Classification

Data were used to train the neural network. In each training stage the network performance was assessed using testing data set and the best performed network was taken for the whole data set simulation. Best performed network requires correct number of neurons in the hidden layer at the time of training.

- ❖ The number of hidden neurons should be between the size of input layer and the size of output layer.
- ❖ The number of hidden neurons should be 2/3 of input layer size plus the size of output layer.
- ❖ The number of hidden neurons should be less than the twice of input layer size.

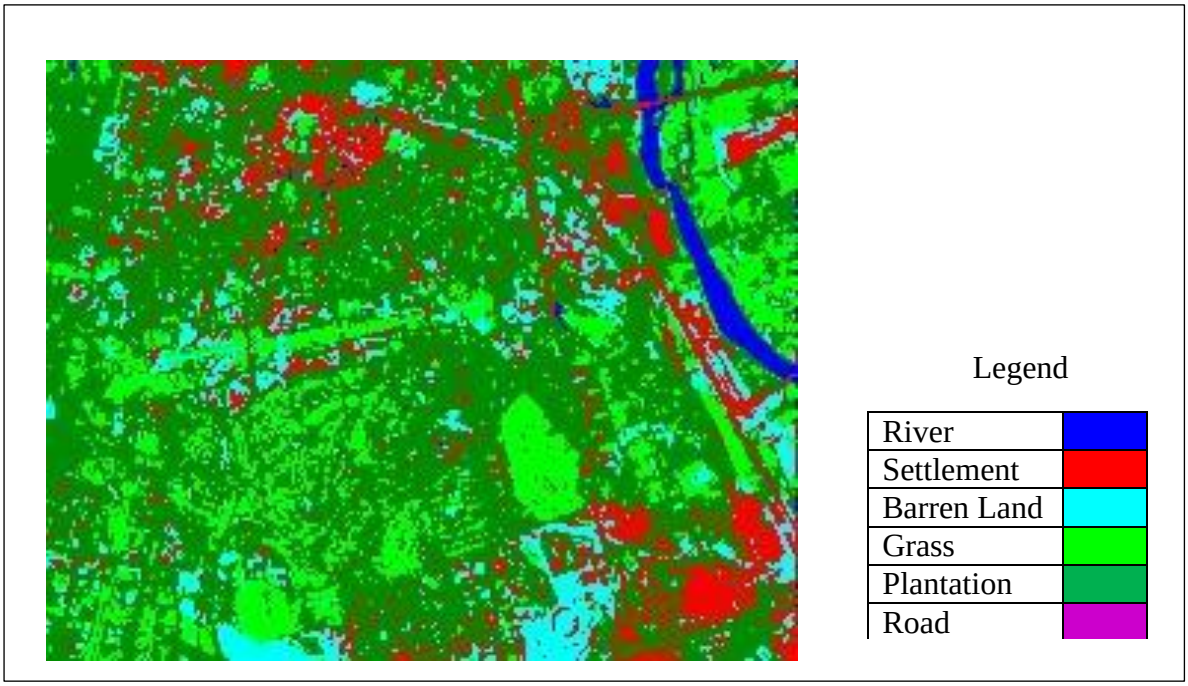

Figure 3: Artificial Neural Network Classification Map

3.2 Unsupervised Classification

Today several different unsupervised classification algorithms are commonly used in remote sensing. The two most frequently used algorithms are the K-mean and the ISODATA clustering algorithm.

Both of these algorithms are iterative procedures. In general, both of them assign first an arbitrary initial cluster vector. The second step classifies each pixel to the closest cluster. In the third step the new cluster mean vectors are calculated based on all the pixels in one

cluster. The second and third steps are repeated until the "change" between the iteration is small. The "change" can be defined in several different ways; either by measuring the distances the mean cluster vector has changed from one iteration to another or by the percentage of pixels that have changed between iterations.
The ISODATA algorithm has some further refinements by splitting and merging of clusters. Clusters are merged if either the number of members (pixel) in a cluster is less than a certain threshold or if the centers of two clusters are closer than a certain threshold. Clusters are split into two different clusters if the cluster standard deviation exceeds a predefined value and the number of members (pixels) is twice the threshold for the minimum number of members.
The ISODATA algorithm is similar to the k-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the k-means assumes that the number of clusters is known a priori.

3.2.1 The ISODATA Unsupervised Classification

This drawback can be overcome in the *ISODATA Algorithm* (Iterative Self-Organizing Data Analysis Technique Algorithm), which allows the number of clusters to be adjusted automatically during the iteration by merging similar clusters and splitting clusters with large standard deviations.

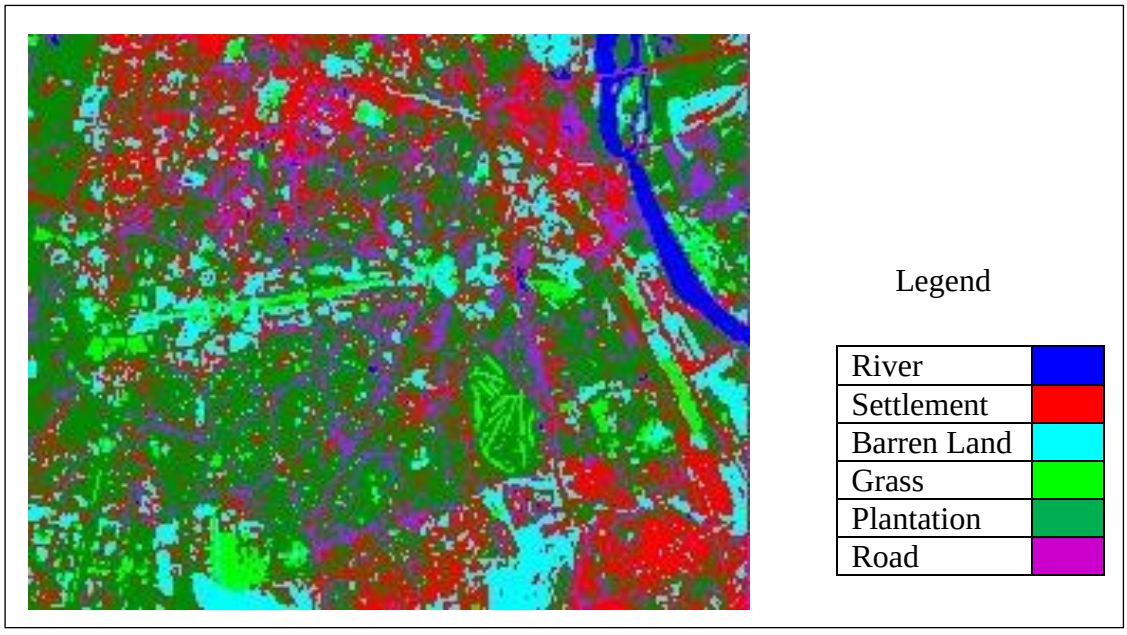

Figure 4: The ISODATA Unsupervised Classification Map

3.2.2 K-means Unsupervised Classification

A. A set number of cluster centres are positioned randomly through the spectral space.
B. Pixels are assigned to their nearest cluster.
C. The mean location is re-calculated for each cluster.
D. Repeat 2 and 3 until movement of cluster centres is below threshold.
E. Assign class types to spectral clusters.

In the K-means method, the number of clusters K remains the same throughout the iteration, although it may turn out later that more or fewer clusters would fit the data better.

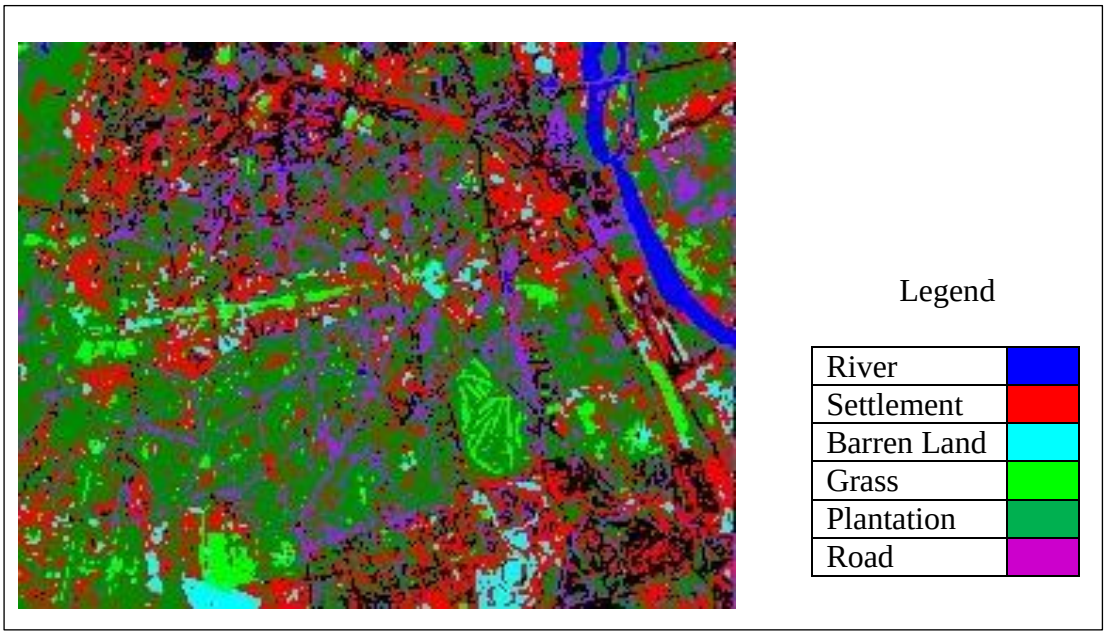

Figure 5: K-means Unsupervised Classification Map

3.3 Normalized Difference Vegetation Index (NDVI)

The pixel values of the NDVI data layer range from -1 to+1 and are scaled from 0 to 255 respectively. The positive values represent different types of vegetation classes, whereas near zero and negative values indicate non-vegetation classes, such as water, snow and barren land.

The surface was covered with many different features which include rocks, vegetation cover, water body and roads. Large area was covered with vegetation class. It was necessary to mask out the vegetation areas. For this purpose NDVI was calculated and vegetation was delineated. Generalized formula for NDVI is as follows:

$$NDVI = \frac{NIR - R}{NIR + R}$$

For calculating NDVI for Hyperion this formula was transformed as follows:

$$NDVI = \frac{\lambda^{925.404} - \lambda^{650.6727}}{\lambda^{925.404} + \lambda^{650.6727}}$$

Vegetation was classified into three broad categories i.e sparse, moderate and dense. The area under study is covered and divided with no vegetation, sparse, moderate and dense vegetation as shown in the Figure - 6 Colour code of the study area of classes versus NDVI values is shown in Table 1.

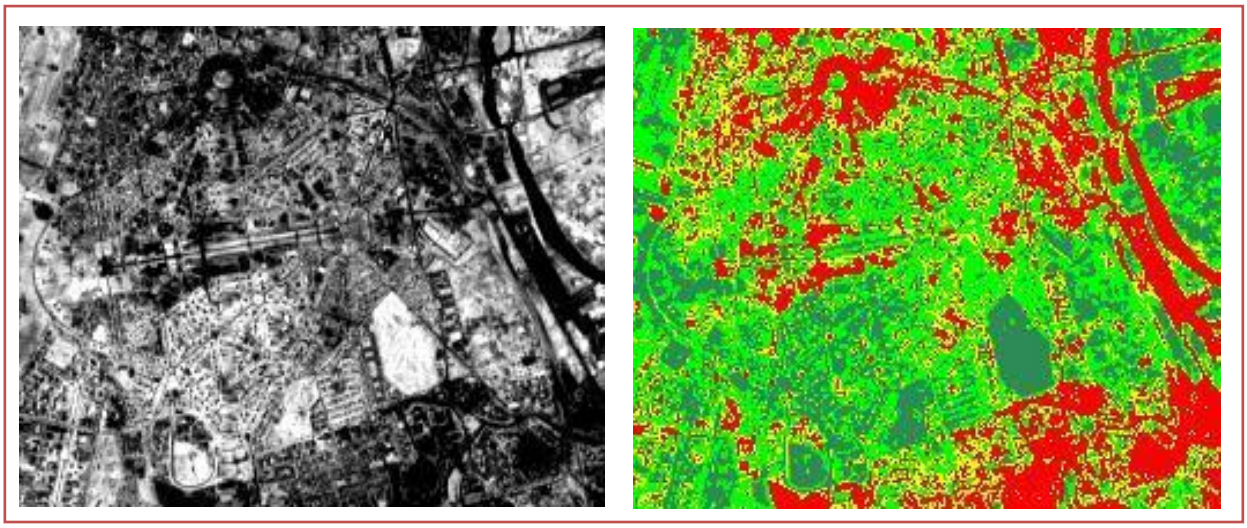

Figure 6: Normalized Difference Vegetation Index (NDVI) image from Hyperion Hyperspectral imagery.

Table 1: NDVI Vegetation Delineation

S. No.	Classes	NDVI	Color Code
1	No Vegetation	-1 to 0.12	
2	Sparse Vegetation	0.12 to 0.24	
3	Moderate Vegetation	0.24 to 0.50	
4	Dense Vegetation	0.50 to 1.00	

4. Accuracy Assessment results

Accuracy assessment should be an important part of any classification. I say "should" because it is frequently not done. The reason for this is that it usually involves a lot of work in the field, which can be very expensive and time consuming. However, without any accuracy assessment, we do not know how accurate our classification is: Are we right about 90% of the time with our classification, or maybe only 50%?

The accuracy of a classification is usually assessed by comparing the classification with some reference data that is believed to accurately reflect the true land-cover. Sources of reference data include, among other things ground truth, higher resolution satellite images, and maps derived from LISS - IV. Note that virtually all reference data (even ground truth data) are inaccurate to some degree as well.

Also, accuracy assessment should not be based on the training pixels. The problem with using training pixels is that they are usually not randomly selected and that the classification is not independent of the training pixels. Using training pixels usually results in an overly optimistic accuracy assessment. The results of an accuracy assessment are usually summarized in a confusion matrix:

Table 2: Validation results obtained from SVM classified image

Accuracy report (Confusion Matrix)								
		Ground Truth						No. of Classified Pixels
	Class Name	River	Grass	Plantation	Settlement	Road	Barren Land	
Classified Image	River	**40**	0	0	0	4	0	44
	Grass	0	**20**	4	0	0	1	25
	Plantation	0	7	**62**	1	1	1	72
	Settlement	0	1	1	**52**	5	1	60
	Road	0	0	0	2	**15**	1	18
	Barren Land	0	1	0	1	1	**29**	32
No. of Ground Truth Pixels		40	29	67	56	26	33	**251**

The user and producer accuracy are two widely used measures of class accuracy. The producer's accuracy refers to the probability that a certain land-cover of an area on the ground is classified as such, while the user's accuracy refers to the probability that a pixel labeled as a certain land-cover class in the map is really this class. The user and producer accuracy for any given class typically are not the same. In the above example, an estimate for the producer accuracy of Plantation is:

$$\text{Producer Accuracy } (PA) = \frac{62}{67} * 100 = \mathbf{92.54\%}$$

While the user accuracy is:

$$\text{User Accuracy } (UA) = \frac{62}{72} * 100 = \mathbf{86.11\%}$$

As a user of a classification, I can expect that roughly 86.11% of all the pixels classified as Plantation are indeed Plantation on the ground. However, as a producer, I am quite unsettled by the fact that I only classified 92.54% of all the Plantation pixels as such. All feature producer accuracy and user accuracy results are shown Table 3.

While the overall accuracy and kappa value is:-

$$\text{Overall Accuracy of classification is } (OA) = \frac{40+20+62+52+15+29}{251}$$

$$= \frac{218}{251} = 0.8685 * 100 = 86.85\ \%$$

$$\text{The kappa value } k \text{ is } = \frac{(218x251)-[(44x40)+(25x29)+(72x67)+(60x56)+(18x26)+(32x33)]}{251^2-[(44x40)+(25x29)+(72x67)+(60x56)+(18x26)+(32x33)]}$$

$$= \frac{42525}{50808} = 0.8370 \qquad \boldsymbol{k} = \mathbf{0.8370}$$

The accuracy assessment is performed by calculating confusion matrix between the ground truth data and the classified image. This process is performed on every the classified image. The results of the SVM, SAM, ANN, ISODATA and K-means classified image are given in the table 3. We are showing here accuracy assessment method of SVM classified image. We are presenting hear User accuracy, Producer accuracy, Overall accuracy and Kappa value of every classified image.

SVM classified image result is best among all classified image. The overall accuracy of the SVM classified image is 86.85 % and kappa value is 0.8370.

Table 3: Producer Accuracy, User Accuracy results obtained from SVM, SAM, ANN, ISODATA, and K-means classified image

Accuracy report of SVM, SAM, ANN, ISODATA and K-MEANS (Confusion Matrix)																				
Class	SVM				SAM				ANN				ISODATA				K-Means			
	User Accuracy	Producer Accuracy	Overall Accuracy	Kappa	User Accuracy	Producer Accuracy	Overall Accuracy	Kappa	User Accuracy	Producer Accuracy	Overall Accuracy	Kappa	User Accuracy	Producer Accuracy	Overall Accuracy	Kappa	User Accuracy	Producer Accuracy	Overall Accuracy	Kappa
River	90.91 %	100.00 %	**86.85 %**	**0.8370**	91 %	100 %	79.68 %	0.7508	86.36 %	100 %	74.90 %	0.6881	90.91 %	100 %	82.07 %	0.7787	88.64 %	100 %	79.68 %	0.7488
Grass	80.00 %	68.97 %			72.00 %	62.07 %			60.00 %	44.12 %			72.00 %	52.94 %			68.00 %	53.13 %		
Plantation	86.11 %	92.54 %			79.17 %	90.48 %			83.33 %	84.51 %			80.56 %	92.06 %			81.94 %	89.39 %		
Settlement	86.67 %	92.86 %			80.00 %	90.57 %			73.33 %	83.02 %			80.00 %	87.27 %			78.33 %	87.04 %		
Road	83.33 %	57.69 %			66.67 %	50.00 %			55.56 %	50.00 %			77.78 %	51.85 %			66.67 %	42.86 %		
Barren Land	90.63 %	87.88 %			78.13 %	80.65 %			65.63 %	60.00 %			87.50 %	87.50 %			81.25 %	81.25 %		

5. Conclusions

In conclusion, the proposed classification methodology succeeding taking advantage of the spatial and the spectral information simultaneously. The method performs well for images representing different scenes: those containing large spatial structures with spectrally confusing classes and those containing small and complex structures. Furthermore, its efficient implementation is possible.

NDVI and vegetation delineation is used to for vegetation classes feature extraction. Spectral profiles are used for feature extraction of water and build-up areas of the study area.

Remote sensing is becoming promising research field in current days. Hyperspectral classification become promising task because of some challenging task like Curse of

dimensionality, Few labeled samples, the spatial variability of the spectral signature, exploring Spatial correlation among pixels and adding contextual information along with spectral information during classification. This letter gives some well-known techniques based on how training samples are used to classify the hyperspectral data. Future research will be in direction of constructing hypergraph and conducting learning on hypergraph to get final classification results.

REFERENCES

1. *Wang Rosalind Xuyan, aug 2008. Learning and Classification of Hyperspectral Images, thesis*
2. *Plaza A. , Benediktsson J.A. , Boardman J.W., Brazile J., Bruzzone L., Camps-Valls G. and Trianni. G., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment, 113:110{122, 2009}*
3. *Chang C.I., 2003. Hyperspectral imaging: techniques for spectral detection and classification. Kluwer Academic/Plenum Publishers: New York.*
4. *Chang C.I., 2007. Recent advances in hyperspectral signal and image processing, John Wiley & Sons: New York.*
5. *Bioucas-Dias J.M., Plaza A., Dobigeon N., Parente M., Du Q., Gader P. and Chanussot J., 2012. Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(2):354{379, 2012}.*
6. *Campbell, J. B, 1996. Introduction to remote sensing (2nd ed.). London: Taylor and Francis.*
7. *Plaza, J., Plaza, A., Perez, R. and Martinez, P., 2005. Automated generation of semi labeled training samples for nonlinear neural network-based abundance estimation in hyperspectral data, IGARSS 2005, Seoul (S. Korea), pp. 1261–1264.*
8. *Plaza, J., Plaza, A., Perez, R. and Martinez, P., 2009.On the use of small training sets for neural network-based characterisation of mixed pixels in remotely sensed hyperspectral images, Pattern Recognition, 42, pp. 3032–3045.*
9. *Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis, International Journal of Remote Sensing, 22, pp. 1305–1334.*
10. *Du, P., Tan, K. and Xing, X., 2010.Wavelet SVM in reproducing kernel Hilbert space for hyperspectral remote sensing image classification. Optics Communications, 283, pp. 4978–4984.*
11. *Pignatti, S., Cavalli, R. M., Cuomo, V., Fusilli, V., Pascucci, S., Poscolieri, M., et al., 2009. Evaluating Hyperion capability for land cover mapping in a fragmented ecosystem: Pollino National Park, Italy, Remote Sensing of Environment, 113, pp. 622–634.*
12. *Wang, J., Chen, Y., He, T., Lv, C. and Liu, A., 2010.Application of geographic image cognition approach in land type classification using Hyperion image: A case study in China. International Journal of Applied Earth Observation and Geoinformaiton, 12S, pp. S212–S222.*
13. *Pervez W. , Khan S.A. and Valiuddin, 2015. Hyperspectral hyperion imagery analysis and its application using spectral analysis.*

14. *Fauvel M., Tarabalka Y., Benediktsson J.A., Chanussot J. and Tilton J.C., 2013. Advances in Spectral-Spatial Classification of Hyperspectral Images, Proc. IEEE, Mar. 2013 ,vol. 101, no. 3, pp. 652–675.*
15. *Guo B., Gunn S.R., Damper R.I. and Nelson J.D.B., Oct. 2006. Band Selection for Hyperspectral Image Classification using Mutual Information, IEEE Geosci. Remote Sens.Lett.vol.3, no.4, pp. 522–526.*
16. *Bernard K., Tarabalka Y., Angulo J., Chanussot J. and Benediktsson A., 2012. Spectral-Spatial Classification of Hyperspectral Data Based on a Stochastic Minimum Spanning Forest Approach, IEEE Trans. Image Process, vol. 21, no. 4, pp. 2008–2021.*
17. *Bilgin G., Erturk S. and Yildirim T., Oct 2008. Unsupervised Classification of Hyperspectral-Image Data using Fuzzy Approaches that Spatially Exploit Membership Relations, IEEE Geosci. Remote Sens. Lett. vol. 5, no. 4, pp. 673–677.*
18. *Huang Y., Liu Q., Lv F., Gong Y., and Metaxas D., Jun 2011. Unsupervised Image Categorization by Hypergraph Partition, IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 6, pp. 1266–1273.*
19. *Sabale Savita P. and Chhaya R. Jadhav, 2014. Supervised, Unsupervised, and Semisupervised Classification Methods for Hyperspectral Image Classification-A Review.*
20. *Boser, B. E., Guyon, I. M. and Vapnik V. N., July 1992. A training algorithm for optimal margin classifiers. In the Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, 27-29 July 1992, Pittsburgh, PA, USA.*
21. *Vapnik and V. N., 1998. Statistical learning theory. New York: Wiley.*
22. *Folkman, M., Pearlman, J., Liao, L., Jarecke and P., 2001. EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration," Proceedings of SPIE 4151, Hyperspectral Remote Sensing of the Land and Atmosphere: 40-51.*
23. *USGS (U.S. Geological Survey), 2011. Ed., "Website Earth Observing 1 (EO-1)," http://eo1.usgs.gov/(10.11.2012).*
24. *Barry and P.S., 2001. [EO-1/ Hyperion Science Data User's Guide, Level 1_B], TRW Space, Defense & Information Systems, Redondo Beach, CA.*
25. *Scheffler Daniel and Karrasch Pieree, 2013. Preprocessing of Hyperspectral Images -a Comparative Study of Destriping Algorithms for EO-1 Hyperion.*
26. *Thenkabail PS, Enclona EA, Ashton MS and Meer B Van Der, 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sensing Envirmenl 91(3-4), pp. 354–376.*
27. *EO1User Guide (2003), Version 2.3.*
28. *ENVI User Guide: Advanced Hyperspectral Analysis.*
29. *EO-1 User Guide, 2003. USGS Earth Resources Observation System Data Centre (EDC).*
30. *Kumar, L., Schmidt, K., Dury, S., Skidmore and A.K, 2001. Imaging spectrometry and vegetation science. In: van de Meer, F., de Jong, S.M. (Eds.), Imaging Spectrometry. Kluwer Academic Press, Dordrecht, pp 111155.*
31. *Prasad, P. R. C., Reddy, C. S., Rajasekhar, G., Dutt and C.B.S, 1992. mapping and Analyzing Vegetation Types of North Andaman Islands, India. GIS, pp.13. [available online] http://www.gisdevelopment.net/application /nrm /overview /over004_1.htm*
32. *Chakravortty Somdatta and Chakrabarti S., 2011. Preprocessing of Hyperspectral Data: A case study of Henry and LothianIslands in Sunderban Region, West Bengal, India*
33. *Dhumal Rajesh K., Rajendra Yogesh, Kale K.V. and Mehrotra S.C., 2013. Classification of Crops from remotely sensed Images: An Overview*

34. *Atkinson P.M. and Tatnall A.R.L., 1997. Introduction to neural networks in remote sensing, International Journal of Remote sensing, volume 11, pp. 699-709.*
35. *Crammer, K. and Singer, Y., 2002. On the Learnability and Design of Output Codes for Multiclass Problems. Machine Learning 47: 201–233.*
36. *Jun Li, 2011. Discriminative Image Segmentation: Applications to Hyperspectral Data, thesis.*
37. *Miguel Plaza Antonio, 2013. New Techniques for Hyperspectral Image Classification, thesis.*
38. *http://www.google.co.in/img*
39. *http://www.google.co.in/introduction+of+delhi*