19th Esri India User Conference 2018

"COMPARISION OF CLASSIFICATION METHODS ON REMOTE SENSED SATELLITE DATA: AN APPLICATION IN GAUTAM BUDDHA NAGAR, INDIA"

Shivangi S. Somvanshi [1], Oshin Bhalla[2]

[1] Assistant Professor, Amity University

[2] Student, Amity University

Sector – 125, Noida

Abstract:

This paper presents classification of various land cover types from the satellite image of Landsat OLI & TIRS obtained from U.S. Geological Survey (USGS), using different classifiers and performances of the classifiers are analyzed. Prior to classification, Training process to assemble a set of statistics describing spectral response pattern of each land cover type is done. The quality of training plays a crucial role in success of classification. Supervised classification is executed based on the spectral features using The Maximum Likelihood Classifier, Minimum Distance Classifier and Mahalanobis Classifier using ARC GIS. Efficiency of Classification results are assessed by using accuracy assessment and Confusion matrix. Performance of Mahalanobis supervised classifier is found to be better than others.

Keywords: Maximum Likelihood Classifier, Minimum Distance Classifier, Mahalanobis Classifier, Accuracy assessment, confusion matrix.

About the Author:

Mrs. Shivangi S. Somvanshi

Mrs. Shivangi Somvanshi is M.Tech in Environmental Engineering and working in the field of Remote Sensing from past 8 Years.

E mail ID: ssomvanshi@amity.du

Contact: +919650384504

Introduction

Xiuwan *et al*. (1999) attempted land cover classification by combining unsupervised algorithm and training data. Steele and Redmond (2001) applied a method for improving land cover mapping. The digital classification techniques were well demonstrated by Lillesand and Kiefer (2004).

B'ardossy and Sammaniego (2002) have applied Fuzzy rule-based classification on remotely sensed digital image data. Lucas *et al*. (2007) used rule-based classification of multitemporal satellite imagery for habitat and agricultural land cover mapping. Ji (2002) did Fuzzy modeling and classification for realistic representation of vegetative characteristic. Kalita and Devi (2002) have used fuzzy supervised classification of remotely sensed image in the presence of untrained classes using Neural Network. Tapia *et al*. (2005) have optimized sampling schemes for vegetation mapping using fuzzy classification. Shalan *et al*. (2003) have reported fuzzy classification to depict a more appropriate land cover in an area where classes were generally mixed.

Neural Network classification has been applied by Augusteijn and Folkert (2002) for identification of ground cover. Murthy *et al*. (2002) have applied Artificial Neural Network and maximum likelihood algorithm for classification of paddy on multi-temporal digital images. Artificial Neural Network for hierarchical classification of multi-date WiFS images was also conducted by Kannan *et al*. (2002). Chitroub (2005) executed a neural network model that was performed on standard PCA and its variants directly from the original remote sensing data. Benediktsson *et al*. (1990) applied neural network approaches verses statistical method in classification of remote sensing data. Bischof *et al*. (1992) have used neural network classification on multi spectral Landsat images. Chlorophyll-A concentration was estimated using an artificial neural network based algorithm with Oceansat-1OCM data by Nagamani *et al*. (2007).

Mannan and Roy (2003) have been demonstrated application of crisp and fuzzy algorithms to the classification of multi-spectral IRS image and found fuzzy method showing accurate results in comparison to other. Foody (2005) has derived thematic classification accuracy through confusion matrices. Hrol and Akdeniz (2005) made an application based on Mixture Distribution models for the pre-field classification method and assessed classification accuracy. Janseen *et al*. (1999) presented vegetation monitoring on reliability aspects with high-resolution remote sensing images. Accuracy assessment through error matrix analysis was done by Sharma and Bren (2005). Hixon *et al*. (1980) evaluated several schemes for classification of remotely sensed data. Land use and land cover change, as one of the main driving forces of global environmental changes, is central to the sustainable development debate. Kant *et al*. (1997) assessed micro level landuse changes using remote sensing data.

Changes in land cover by land use do not necessarily imply degradation of the land. However, many shifting land use patterns driven by a variety of social causes, result in land cover changes that affects biodiversity, water, trace gas emission and other processes that come together to affect climate and biosphere (Riebsame et. al, 1994).

According to Olorunfemi (1983), monitoring changes and time series analysis is quite difficult with traditional method of surveying. In recent years, satellite remote sensing techniques have been developed, which have proved to be of immense value for preparing accurate land use, land cover maps and monitoring changes at regular interval of time. In case of inaccessible region this technique is perhaps the only method of obtaining the required data on a cost and time effective basis.

Applications of remote sensing and GIS for change analysis of landuse/ landcover have been done by Xiuwan (2002). Tirkey *et al*. (2005) demonstrated use of remote sensing and GIS to track temporal changes in Mumbai coastal area. Land damage assessment and change detection analysis have been and GIS offers a wide scope in detection of land cover changes over period of time. Zhang et al (2014) used remote sensing data in the national landuse change program of China. Digital change detection techniques based on multi-temporal and multi-spectral remotely sensed data have demonstrated a great potential as a means to understanding landscape dynamics – detect, identify, map and monitor differences in land use and land cover patterns over time, irrespective of casual factors (Jensen, 1996). Recent improvements in satellite image quality and availability have made it possible to perform image analysis at much larger scale than in the past. Change detection is an important process in monitoring and managing natural resources and urban development because it provides quantitative analysis of the spatial distribution of the population of interest. Macleod and Congation (1998) listed four aspects of change detection which are important when monitoring natural resources: (i) Detecting the changes that have occurred (ii) Identifying the nature of the change (iii) Measuring the area extent of the change (iv) Assessing the spatial pattern of the change.

A wide variety of digital change detection techniques have been developed over the last two decades. Coppin & Bauer (1996) summarize eleven different change detection, these include Mono temporal change delineation, Delta or post classification comparisons, Multi-dimentional temporal feature space analysis, Composite analysis, Image differencing, Multi-temporal linear data transformation, Change vector analysis, Image regression, Multi-temporal biomass index, Background subtraction and Image rationing. Pandy and Nathawat in 2006 inferred that land use land cover pattern in the area are generally controlled by agro – climatic conditions, ground water potential and other geological factors, which are very important in developing a land resource management scheme.

Urban land use change monitoring compared, using high-resolution remote sensing technology to monitor more efficient time saving, saving a lot of manpower, material resources and time, improve the urban land use database building and database and update efficiency. The growth of city without planning will lead to create many complex urban problems. Basic amenities such as water, electricity, sewage etc. in this context. Hence efficient classification technique is required.

Material and Methods

(i) Data Sets:

Landsat satellite images for the year 2001 and 2016 was obtained from United States Geology Survey (USGS) and was utilized in this paper. Detailed information about the remotely sensed images is listed in Table 1. In image acquisition, we have considered to the impacts of Sun's inclination, season and cloud cover. In supporting the study, secondary data were obtained from the different organizations like, the city boundaries map, districts shape file and land use maps. We also have conducted fields' observations in study area as a basis for accuracy assessment.

Table 1: Satellite Data used for the study

SATELLITES	ACQUISITION DATE	SENSOR	SPATIAL RESOLUTION	PROJECTION
Landsat 8	02/03/2016	OLI-TIRS	30 m	WG84 UTM Zone 44N
Landsat 5	05/02/2001	TM	30 m	WG84 UTM Zone 44N

(ii) Image Classification:

All the enhanced images were then subjected to image classification. Image classification is the process of extracting the information classes from a multiband raster image. The supervised classification was employed to process and classify the images of the study area. Based on different bands of the Landsat images a composite image was created that helped us to clearly distinguish different classes in an image. The Maximum Likelihood Classifier, Minimum Distance Classifier and Mahalanobis Classifier were used for spectral classification of the Landsat images. Five land cover classes were identified in the study area, namely, urban built-up, rural built-up, wasteland, agricultural land and water

1. Minimum Distance to Means Classifier

This classifier is the simplest classifier. In this process, only the mean vector of each class is used. Other data, such as standard deviations and covariance matrices are ignored. The minimum distance classifier is used to classify unknown image data to classes which minimize the distance between the image data and the class in multi feature space. The distance in Equation 1 is called index of similarity.

$$dk2 = (X-\mu k)\ T\ (X-\mu k) \text{ ----------Eqn (1)}$$

Where X is vector of image data μk is the mean vector of the kth class. It is same as the Euclidean distance between two vectors. The minimum distance classifier is not complete, as it does not take correlations with in dataset.

2. Maximum Likelihood Classifier (MLC)

The Maximum Likelihood Classifier quantitatively evaluates both the varianceand covariance of the category spectral response patterns when classifying an unknown pixel. To do this, an assumption is made that the distribution of the cloud of points forming the category training data is Gaussian (normal distribution). This assumption is generally reasonable for common spectral response distributions. Under the assumption, the distribution manner of a category response pattern can be completely described by the mean vector and the covariance matrix. Given these parameters, we may compute the statistical probability of a given pixel value being a member of a particular land cover class. Figure 4 shows the probability values plotted in a three dimensional graph. The probability density functions (Eqn. 2) are used to classify an unidentified pixel by computing the probability of the pixel value belonging to each category.

$$\hat{p}(x \mid w_i) = \frac{1}{(2\pi)^{\frac{1}{2}}\hat{\sigma}_i} \exp\left[-\frac{1}{2} \frac{(x-\hat{\mu}_i)^2}{\hat{\sigma}_i^{\,2}} \right]$$

………… Equ (2)

The probability density functions (Eq. 2) are used to classify an unidentified pixel by computing the probability of the pixel value belonging to each category. That is, the computer would calculate the probability of the pixel value occurring in the distribution of class “corn”, then the likelihood of its occurring in class “sand”, and so on. After evaluating the probability in each category, the pixel would be given to the most likely class or be labelled “unknown” if the probability values are all below a threshold set by the analyst.

3. Mahalanobis Classifier

In mathematical terms, the Mahalanobis distance is equal to the Euclidean distance when the covariance matrix is the unit matrix. This is exactly the case then if the two columns of the standardized data matrix are orthogonal. The Mahalanobis distance depends on the covariance matrix of the attribute and also accounts for the correlations. To say exactly, the covariance matrix is utilized to correct the effects of cross-covariance between two components of random variable.

The Mahalanobis distance is the distance between an observation and the center for each group in m-dimensional space defined by m variables and their covariance. Thus, a small value of Mahalanobis distance increases the chance of an observation to be closer to the group’s center and the more likely it is to be assigned to that group.

For each feature vector, the Mahalanobis distances towards class means are calculated as in Equation 3. This includes the calculation of the variance-covariance matrix V for each class.

dk2 = (X-µk) TV (X-µk) ----------Eqn (3)

Where X is vector of image data µkis the mean vector of the kth class **V**is the variance covariance matrix

(iii) Accuracy Assessment:

Accuracy assessment for individual classification is essential for an efficient analysis of LULC change (Butt et al., 2015). It indicates the degree of deference between classified images and reference data. 500 random points over each type of classified image was created and analyzed for accuracy, with the help of ground truth data. Based on error matrix the accuracy assessment of LULC maps was carried out. The error matrix compares the relationship between known ground or reference data and the corresponding outcomes of an automated classification.

Accordingly, the commission and commission errors could also be estimated to give percentage of accuracy each class. Accuracy assessment through error matrix analysis was done by Sharma and Bren (2005). Foody (2005) has derived thematic classification accuracy through confusion matrices. Macleod and Congalton (1998) demonstrated that error matrix is the standard way of presenting results of an accuracy assessment. It is a square array in which accuracy assessment sites are tallied by both their classified category in the image and their actual category according to the reference data.

Kappa accuracy is determined from the error matrix, which not only gives the number of correctly classified units but also the errors of commission and omission. An error of omission occurs by erroneously excluding a unit from a category, when it belongs to the same class. An error of commission occurs by erroneously including a unit belonging to some other category. 'Kappa statistics’ (Lillesand *et al.*, 2004) is defined as follows:-

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \quad \text{………… Equ 4}$$

Where
r = number of rows in the error matrix
x_{ii} = number of observations in row *i* and column *i* (on the major diagonal)
x_{i+} = total observations in row *i* (shown as marginal total to right of the matrix)
x_{+i} = total observations in column *i* (shown as marginal total at bottom of the matrix)
N = total number of observations included in matrix

Result and Discussion

Enhanced images using the above mentioned technique were subjected to supervised and unsupervised classification. Error matrices were employed to assess classification accuracies and are detailed for years 2016 in Tables 2–4 using maximum likelihood classifier, minimum distance classifier and mahalanobis classifier as the classification method of supervised and unsupervised classification respectively. The overall accuracy indicates the percentage of correctly classified pixels. The overall accuracies of the classification methods for 2016 ranged from 77.41% to 92.4%, with Kappa coefficients from 0.59 to 0.883. According to Anderson (1976), 85%, as a minimum accuracy value is acceptable. Hence, the accuracy assessment is reliable. After comparing the accuracy statistics of all the three types of classification, mahalanobis classification scheme was detected as the best and most accurate classification method and further used to classify the images of the study area for 2001.

Table 2: Error Matrix comparing the image classification of 2016 to the reference data using maximum likelihood classifier (supervised classification).

Classes	Wasteland	Water body	Agriculture land	Rural built up	Urban built up	Row Total	User's Accuracy (%)
Waste land	**2**	0	1	1	0	**4**	**50**
Water body	0	**12**	1	5	0	**18**	**66.66**
Agriculture land	1	2	**106**	7	0	**116**	**91.38**
Rural built up	1	1	1	**23**	2	**28**	**82.14**
Urban built up	0	4	0	0	**63**	**67**	**94.03**
Column Total	**4**	**19**	**109**	**36**	**65**	**233**	
Producer's Accuracy (%)	**50**	**66.66**	**88.33**	**63.89**	**96.92**	**365.8**	
Overall Accuracy (%)	**92.4**						
Kappa Coefficient	**0.881**						

Table 3: Error Matrix comparing the image classification of 2016 to the reference data using minimum distance classifier (supervised classification).

Classes	Wasteland	Water body	Agriculture land	Rural built up	Urban built up	Row Total	User's Accuracy (%)
Waste land	**3**	4	0	3	2	**12**	**25**
Water body	1	**3**	0	0	0	**4**	**75**
Agriculture land	0	2	**124**	6	0	**132**	**93.23**

Rural built up	1	1	12	**33**	16	**63**	**52.38**
Urban built up	0	0	0	2	**34**	**36**	**94.44**
Column Total	**5**	**10**	**136**	**44**	**52**	**247**	
Producer's Accuracy (%)	**60**	**30**	**91.18**	**75**	**65.38**		
Overall Accuracy (%)	**89.8**						
Kappa Coefficient	**0.844**						

Table 4: Error Matrix comparing the image classification of 2016 to the reference data using mahalanobis classifier (supervised classification).

Classes	Wasteland	Water body	Agriculture land	Rural built up	Urban built up	Row Total	User's Accuracy (%)
Waste land	**1**	0	0	0	0	**1**	**100**
Water body	1	**6**	12	7	0	**26**	**23.08**
Agriculture land	0	3	**88**	5	0	**96**	**91.67**
Rural built up	0	1	3	**36**	3	**43**	**83.72**
Urban built up	0	3	0	0	**68**	**71**	**95.77**
Column Total	**2**	**13**	**100**	**48**	**71**	**234**	
Producer's Accuracy (%)	**50**	**46.15**	**85.44**	**75**	**95.77**		
Overall Accuracy (%)	**92.4**						
Kappa Coefficient	**0.883**						

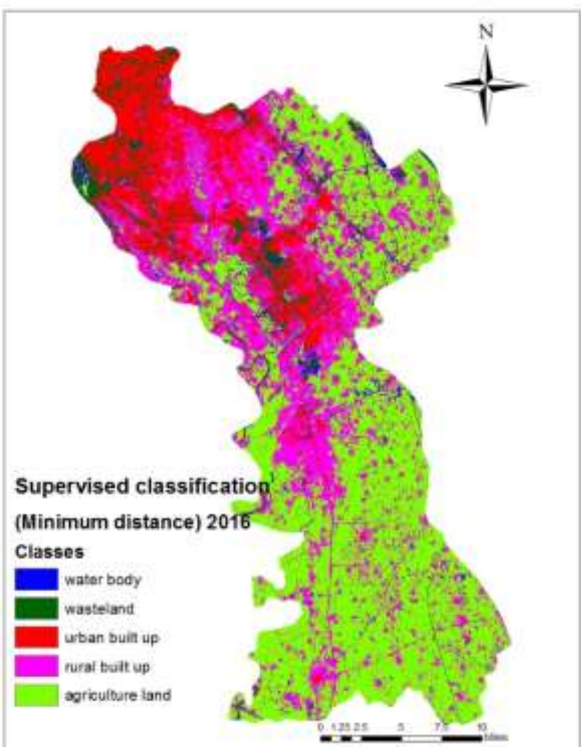

Figure 1: Supervised classification using minimum distance classifier

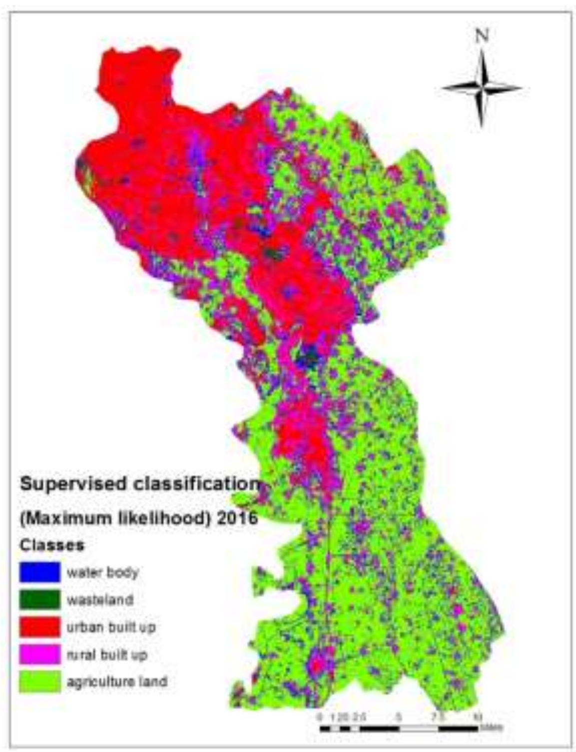

Figure 2: Supervised classification using maximum likelihood classifier

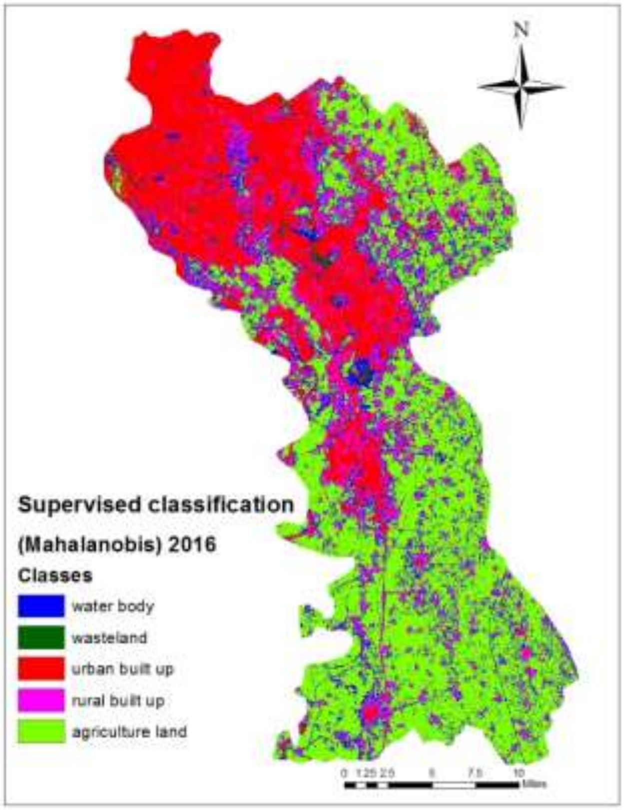

Figure 3: Supervised classification using Mahalanobis classifier

Conclusion

Accurate image classification is the pre requisite for any application of the satellite imagery. Efficiency of Classification results are assessed by using accuracy assessment and confusion matrix. Performance of Mahalanobis supervised classifier is found to be better than others with overall accuracy of 92.4% and kappa coefficient of 0.883.

References

1. Augusteijn, M.F. and Folkert, B.A. (2002). Neural Network Classification and Novelty Detection, *Int. J. Remote Sensing*, 23(14): 2891-2902.
2. Benediktsson, J.A., Swain, P.H., and Ersoy, O.K. (1990). Neural Network Approaches versus Statistical Methods in Classification of Multisource Remote Sensing Data, *IEEE, Transaction on Geoscience and Remote Sensing*, 28(4): 540-552.
3. Bischof, H., Schneider, W. and Pinz, A.J. (1992). Multispectral Classification of Landsat Images Using Neural Networks, *IEEE, Transaction on Geoscience and Remote Sensing*, 30(3): 482-490.
4. Butt, A., Shabbir, R., Ahmad, S.S., Aziz, N., (2015). Land use change mapping and analysis using Remote Sensing and GIS: a case study of Simly watershed, Islamabad, Pakistan. Egypt. J. Remote Sens Space Sci. 18 (2), 251–259.
5. Chitroub, S. (2005). Natural Network Model for Standard PCA and Its Variants Applied to Remote Sensing, *Int. J. Remote Sensing*, 26(10): 2197-2218.
6. Coppin, P. & Bauer, M. (1996). Digital Change Detection in Forest Ecosystems with Remote Sensing Imagery. Remote Sensing Reviews. Vol. 13. p. 207-234.

7. Foody, G.M. (2005). Local Characterization of Thematic Classification Accuracy through Spatially Constrained Confusion Matrices, *Int. J. Remote Sensing*, 26(6): 1217-1228.
8. Foody, G.M. (2005). Local Characterization of Thematic Classification Accuracy through Spatially Constrained Confusion Matrices, *Int. J. Remote Sensing*, 26(6): 1217-1228.
9. Harika D, Swamy AVVS, John Vijay T (2012) Potential of Jatropha curcas on large scale cultivation as a renewable biodiesel alternative. Int J Environ Sci. 3:393-397.
10. Hixon, M., Scholz, D. and Fuhs, N. (1980). Evaluation of Several Schemes for Classification of Remotely Sensed Data, *Photogrammetric Engineering and Remote Sensing*, 58: 423-437
11. Hixon, M., Scholz, D. and Fuhs, N. (1980). Evaluation of Several Schemes for Classification of Remotely Sensed Data, *Photogrammetric Engineering and Remote Sensing*, 58: 423-437
12. Hrol, H. and Akdeniz, F. (2005). A per-field Classification Method Based on Mixture Distribution Models and An Application to Landsat Thematic Mapper Data, *Int. J. Remote Sensing*, 26(6): 1229-1244.
13. Janseen, J.A.M., Kloosterman, E.H., Jordons, R.W.L. Hartmann, R.A. and Jorritsma, J.S. (1999). Reliability Aspects of Vegetation Monitoring with High Resolution Remote Sensing Images, *Proceedings of the 18th EARSeL Symposium on Operational Remote Sensing for Sustainable Development held at Enschede, Netherlands*, from May 11-14, 1998. (Eds: Nieuwenhuis, G.J.A., Vaughan, R.A. and Moleaar, M.) Operational Remote Sensing for Sustainable Development, pp. 53-60.
14. Kannan, J., Acharekar, J. V. and Mohan B. K. (2002). Hierarchical Classification of Multidate WiFS Images using Artificial Networks, *Proceedings of the ISPRS Commission VII Symposium on Resource and Environmental Monitoring* held at NRSA, Hyderabad, India, from December 03-06, 2002. (Eds: Navalgund, R.R., Nayak S.R., Sudarshan, R., Nagaraja, R. and Ravindran, R.), pp 93-95.
15. Kant, Y., Nigam, R.K. and Jayanti, S.C. (1997). Change Detection of Micro Level for Landuse/ Landcover Using Remote Sensing Data, *Symposium of Remote Sensing for Natural Resources with Special Emphasis on Infrastructure Development* held at NRSA, Hyderabad from November 26-28, 1997. A Joint ISRS, Dehradun and NNRMS, Bangalore Publication (Eds: Rabindran, K.V., Prasad, J., Pande, L.M., Kushwaha, S.P.S. and Saha, S.K.) Remote Sensing and Geographical Information System for Natural resources, pp. 65-74
16. Lillesand, T.M., Kiefer, R.W. and Chipman, J.W. (2004). Remote Sensing and Interpretation, John Wiley and Sons Inc.
17. Lucas, R., Rowlands, A., Brown, A., Keyworth, S. and Bunting, P. (2007). Rule-based Classification of Multitemporal Satellite Imagery for Habitat and Agricultural Land Cover Mapping, Int. *J. Photogramentry and Remote Sensing*, 62(3): 165-185.
18. Macleod & Congalton. (1998). A Quantitative Comparison of Change Detection Algorithms for Monitoring Eelgrass from Remotely Sensed Data. Photogrammetric Engineering & Remote Sensing. Vol. 64. No. 3. p. 207 - 216.
19. Maini, R. and Agrawal, H., (2010). A comprehensive review of image enhancement techniques, *Journal of computing*, 2 (3), 8-13.
20. Mannan, B. and Roy, A.K. (2003). Crisp and Fuzzy Competitive Learning Networks for Supervised Classification of Multispectral IRS Scenes, *Int. J., Remote Sensing,* 24(17): 3491 3502.
21. Mannan, B. and Roy, A.K. (2003). Crisp and Fuzzy Competitive Learning Networks for Supervised Classification of Multispectral IRS Scenes, *Int. J., Remote Sensing,* 24(17): 3491 3502.
22. Murthy, C.S., Raju, P.V., Hakeem, K. Abdul and Badrinath K.V.S. (2002). Artificial Neural Network & Maximum Likelihood for Classification of Paddy with Multi-Temporal Reflectance Images, *Proceedings of the ISPRS Commission VII Symposium on Resource and Environmental Monitoring* held at NRSA, Hyderabad, India, from December 03-06, 2002. (Eds: Navalgund, R.R., Nayak S.R., Sudarshan, R., Nagaraja, R. and Ravindran, R.), pp 96 99

23. Nagamani, P.V., Chauhan, P. and Dwivedi, R.M. (2007). Estimation of Chlorophyll-A Concentration Using an Artificial Neural Network (ANN) - Based Algorithm with Oceansat-1OCM Data, *Photonirvachak, J. Indian Society Remote Sensing*, 35(3): 201-207.
24. Olorunfemi J.F (1983). Monitoring Urban Land – Use in Developed Countries – An aerial photographic approach, Environmental Int.9, 27 – 32.
25. Pandy A. C. and Nathawat M. C. (2006). Land Use Land Cover Mapping Through Digital Image Processing of Satellite Data – A case study from Panchkula, Ambala and Yamunanagar Districts, Haryana State, India.
26. Riebsame, W.E., Meyer, W.B., and Turner, B.L. II. (1994). Modeling Land-use and Cover as Part of Global Environmental Change. Climate Change. Vol. 28. p. 45.
27. Shalan, M.A., Arora, M.K. and Ghosh, S.K. (2003). An Evaluation of Fuzzy Classifications from IRS 1C LISS III Imagery: A Case Study, *Int. J. Remote Sensing*, 24(15): 3179-3186.
28. Sharma, D.P. and Bren, L. (2005). Effect of Seasonal Spectral Variation on Land Cover Classification, *Photonirvachak, J. Indian Society Remote Sensing*, 33(2): 204 -209.
29. Sharma, D.P. and Bren, L. (2005). Effect of Seasonal Spectral Variation on Land Cover Classification, *Photonirvachak, J. Indian Society Remote Sensing*, 33(2): 204 -209.
30. Steele, B.M. and Redmond, R.L. (2001). A Method of Exploiting Spatial Information for Improving Classification Rules Application to the Construction of Polygon-based Land Cover Maps, *Int. J. Remote Sensing*, 22(16): 3143-3166.
31. Tapia, R., Steen, A. and Bijker, W. (2005). Optimization of Sampling Schemes for Vegetation Mapping Using Fuzzy Classification, *Int. J. Remote Sensing Environment*, 99 : 425-433
32. Tirkey, N., Biradar, R.S., Pikle, M. and Charatkar, S. (2005). A Study on Shoreline Changes of Mumbai Coast Using Remote Sensing and GIS, *Photonirvachak, J. Indian Society Remote Sensing*, 33(1): 85-91.
33. Xiuwan, C. (2002). Using Remote Sensing and GIS to Analyse Landcover Change and its Impact on Regional Sustainable Development, *Int. J. Remote Sensing*, 23(1): 107-124.
34. Zhang, Y., Mishra, R.K., (2014). From UNB PanSharp to Fuze Go – the success behind the pan-sharpening algorithm. *International Journal of Image and Data Fusion*, 5, 39–53.