ANALYSIS

39(180), September 1, 2015

ISSN 2278–5469 EISSN 2278–5450

Discovery

# Spatial pattern of urban landscape in Delhi NCT and its Peri-urban area using support vector machine technique

**Dutta D[1], Rahman A[2], Paul SK[3]**

1. Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India; Email: iirs.dipa@gmail.com
2. Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India; Email: ateqgeo@yahoo.co.in
3. Department of Architecture & Regional Planning, Indian Institute of Technology, Kharagpur, India; Email: skpaul@arp.iitkgp.ernet.in

**Publication History**

Received: 09 June 2015

Accepted: 02 August 2015

Published: 1 September 2015

**Citation**

Dutta D, Rahman A, Paul SK. Spatial pattern of urban landscape in Delhi NCT and its Peri-urban area using support vector machine technique. *Discovery*, 2015, 39(180), 166-172



## ABSTRACT

The geospatial techniques have played an important role over the last couple of decades for identifying the land use land cover pattern of an area. Unlike the conventional methods of surveying this technique has become popular for its robustness in data analysis, database management and accuracy. The present study aims to identify the spatial pattern of urban landscape of Delhi NCT and its periurban area using Support Vector Machines (SVM). In order to extract the land use land cover information of the study area, Landsat image of 2014 was classified. Since the raw images consist of radiometric and geometric errors, before classification they were pre-processed using the FLAASH atmospheric correction followed by image to image rectification. The accuracy of classified LULC map of 2014 (K^=0.91) was found significantly high while compared with ground based data. The SVM classifier with principal components has proved a close agreement between classified and referenced data. The study reveals that high density built-up area was dominant class in Delhi NCT region whereas agricultural land occupies the dominant position in the periurban area. It was estimated that about 420.59 sq. km. (28.42 %) of the total NCT region was under high density built-up land which, together with the low density built-up area (19.01 %) represents almost half of the total area. Whereas it depicts a different scenario in the periurban area where high density and low density built-up class covers an area of 215.70 (8.25 %) and 558.16 (21.34 %) sq. km. respectively. Unlike the periurban area, Delhi NCT region having many distinct patches of dense vegetation regardless of its high rate of urbanization. The study reveals that there is a great potentiality of SVM technique for studying the spatial pattern of urban landscape.

**Keywords**: urban landscape, landuse and landcover, geospatial techniques, Support Vector Machine.

## 1. INTRODUCTION

Megacities like Delhi are being changed everyday mainly due to anthropogenic activities which causing increasing pressure on land and natural environment. Rapid development of cities without proper planning and ecological concern has been a great challenge to the urban planners as well the policy makers to manage a livable environment for city dwellers. Development of new urban areas and expansion of existing cities is inevitable as it's an essential part of sustainable economy but uncontrolled and haphazard urban growth may raise serious problems related to environmental pollution, changes in urban micro climate, loss of biodiversity and ecological balance, human and traffic congestion and moreover quality of urban life. As land is a limited resource and over-exploitation of land may exaggerate problem of land degradation, optimum land use is required to achieve maximum benefit. In this context, proper planning and management is essentially required for best utilization of land to meet the socio-economic demand as well as to preserve the sustainability.

In order to identify land use land cover of an area, the geospatial techniques has played an important role over the last couple of decades. Unlike the conventional methods of surveying this technique has become popular for its robustness in data analysis, database management and accuracy. It is also noteworthy that acquisition of satellite based data products does not require much time and cost in comparison with conventional field based methods of data collection (Da Costa, 1999). Since remote sensing technique allows acquisition of spatial data in multi-resolution, multi-spectral and multi-temporal form, it has been accepted as an essential tool for mapping and monitoring land use land cover dynamics (Kushwaha et al., 1996).

Actual information on spatial distribution of different land use and land cover has multi-dimensional utility in planning and management of the land resources which is perceived as a key factor in the process of development of an area. However, optimal use of land resource requires quantitative information on spatial distribution as well as spatio-temporal changes of various land use and land cover in an area. Remote sensing technique has become an established and well accepted tool for acquiring information on land dynamics. In order to convert the satellite based data into reliable land use and land cover information several image classification techniques are being used. The last couple of decades have seen development of many image classifier algorithms, extending from conventional per-pixel based parametric algorithms (e.g., maximum likelihood classification) to advanced nonparametric algorithms (e.g., neural network, decision tree classification, and support vector machine), and to object-based algorithms (Lu and Weng, 2007). Although traditional supervised and unsupervised classifiers have been used successfully by several studies, still there was a gap between targeted and actual accuracy of classified images. Throughout the last decades researches have put their efforts towards minimizing the error and increasing the level of accuracy of image classifiers. As a result, a number of new algorithms for image classification (eg. neural network, decision tree classification) have come up and become popular.

Originally, the Support Vector Machines (SVMs) were rooted from Statistical Learning Theory and widely applied in recognition of optical character, handwriting digit and text (Vapnik, 1995; Joachims, 1998). It has been adopted as a robust classifier in the arena of remote sensing due to its unparallel performance in detecting land use and land cover features (Huang et al, 2002; Mahesh and Mather, 2003; Guo et al., 2005; Pal and Mather, 2005). Camps-Valls et al. (2004) found it better than neural networks as training neural and neurofuzzy models is unfeasible when working with high-dimensional input spaces and great amounts of training data. The superior performance of SVM has proved its ability to generalize well even with limited training samples (Mountrakis et al., 2011). The remote sensing implementations of support vector machines (SVMs) is acknowledged as a promising machine learning methodology (Mountrakis, 2011; Nooni et al., 2014) which can be applied for the purpose of mapping urban land cover, using medium spatial resolution imagery (Griffiths et al, 2010; Poursanidis et al., 2015).

Since the data points closest to the hyperplane are used to measure the margin, they are termed as 'support vectors' which are most critical elements of the training set (ENVI 4.8). Although SVMs are intrinsically binary classifiers (Melgani and Bruzzone, 2004) it can applied for classifying multiple land use land cover features. There are two common approaches to overcome this problem, One-Against-One (1A1) and One-Against-All (1AA) techniques. In order to classify multi-class image, the 'one against one' method should be employed (Pal and Mather, 2005).

Monitoring the spatial pattern of urban landscape using geo-spatial techniques has been attempted by classifying satellite based images and quantifying the amount of area under various land use and land cover classes. Addressing the estimation of urban landuse and landcover has been a major topic of research considering its demand in urban planning as well as policies. Satellite based data coupled with geo-information techniques have made this study more accurate, robust without spending much time and cost. In this context, the present study attempts to assess the potentiality of Support Vector Machine (SVM) technique in identifying the pattern of urban landscape in Delhi NCT and its sarrounding regions from multi-spectral satellite image.

## 2. STUDY AREA

The study area comprises National Capital Territory (NCT) of Delhi and its periurban area. The term 'periurban' indicates the rural-urban fringe which experience continuous change as a result of rapid urbanization. Therefore demarcation of the periurban boundary is a difficult job as there is no discrete boundary for identifying the fringe areas. A fifteen kilometer buffer area from the existing administrative boundary of NCT has chosen as periurban area of NCT. This area contains part of Sonepat, Jhajjar, Gurgaon and Faridabad districts of Hariyana, and part of Goutam Buddha Nagar, Ghaziabad and Baghpat districts of Uttar Pradesh. The NCT of Delhi is having a total area of 1483 sq.km whereas the periurban of NCT covers 2618 sq. km. area. Geographically, this area is situated at the heart of the Indian sub- continent extending between 28°16′10′′N to 29°0′52′′N latitude and 76°48′36′′E to 77°30′9′′E longitude.

## 3. DATA AND METHODOLOGY

The Landsat 8, also known as Landsat Data Continuity Mission (LCDM) launched in 11$^{th}$ February, 2013 and continuing the legacy of landsat program to serve the world good quality satellite data free of cost. Unlike the TM or ETM+ series, the landsat 8 provides eleven spectral bands among which first nine bands including one PAN band are of onboard Operational Land Imager (OLI) sensor and remaining are of Thermal Infrared Sensor (TIRS). The landsat 8 operates in the visible, near-infrared, short wave infrared and thermal infrared region of electromagnetic spectrum and having spatial resolution of 30 and 15 meters for multispectral and PAN images respectively. The enhanced radiometric resolution (16 bit) of landsat 8 has made it useful for identifying land surface features accurately. Field verification was carried out during February, 2014 and seventy sample points were collected using a Trimble GPS. The sample points were selected according to the stratified sampling method where each land use land cover class was considered as a stratum and equal number of sample points was collected from all classes.

Atmospheric effects can reduce the information content of an image by increasing the noise in the signal recorded by a sensor. In order to extract actual quantitative information of the earth surface and measuring the changes of dynamic earth in a quantitative way, the noises carried by signals must be removed. The atmospheric correction of the satellite image was done by MODTRAN (Moderate Resolution Atmospheric Radiance and Transmittance Model) based Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) model. The FLAASH model requires radiometrically calibrated radiance image with band-interleaved-by-line (BIL) or band-interleaved-by-pixel (BIP) format. Therefore, the raw input images were corrected using their respective sun elevation angle and actual time of acquisition. Then, pixel values of the raw image were converted into radiance values. The Principal Component Analysis was employed on the images for transforming the original multispectral bands into principal components. It is a useful statistical technique for reduction of dimensionality of data which helps to extract information as much as possible from an image.

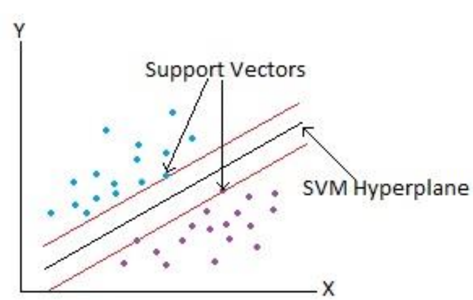


**Figure 1** Classification using support vectors and separating hyperplane (Burges, 1998)

The SVM algorithms has achieved better results than traditional parametric classifiers (Pal and Mather, 2005; Nooni et al., 2014) and even better than nonparametric decision tree classifiers in terms of accuracy, simplicity, and robustness (Foody and Mathur, 2004). The support vector machine separates the classes with a decision surface called the *optimal hyperplane* that maximizes the margin between the classes (Figure 1). Boundary between two classes is estimated by the greatest margin between the two classes, where margin is defined as the sum of the distances to the hyperplane from the closest points of the two classes (Vapnik 1995). The radial basis function (RBF) kernel was used for the present study as it produces best result and gives highest overall classification accuracy (Pal and Mather, 2005). The mathematical representation of this kernel is mentioned in equation 1:

$$\text{RBF} \qquad K(x_i,x_j) = \exp(-g\|x_i - x_j\|^2),\ g > 0$$

.................Equation (1)

where:

g is the gamma term in the kernel function for all kernel types except linear.
$^d$ is the polynomial degree term in the kernel function for the polynomial kernel.
r is the bias term in the kernel function for the polynomial and sigmoid kernels.
g, $^d$, and r are user-controlled parameters, as their correct definition significantly increases the accuracy of the SVM solution.

## 3. RESULT AND DISCUSSION

### 3.1. Land Use and Land Cover of Delhi NCT and its Periurban Area in 2014

It is clearly evident from the classified map (Figure 2) that most of the area within Delhi NCT was under built-up class during the year 2014. Although agricultural land was dominant in periurban areas, large patches of built-up area can be seen in eastern and southern part. It is also observed that numerous small patches of built-up areas are scattered throughout the periurban area which actually indicates growth of small towns and urbanization. A distinct green patch can be identified in the core of the city that represents Delhi ridge forest. Towards the eastern part of the city, some patches of wasteland are observed just adjacent to the low density built-up area. During field survey it is seen that these areas are actually vacant lands being ready for new construction. Earlier, these areas were agricultural land and now it has become converted for the development of new urban centers. It was observed that these areas have experienced a huge real estate boom in recent years followed by mushrooming of high rises and cluster of new apartments. Towards the north eastern part of the periurban area some patches of open spaces can be found. The field verification revealed that these areas are excavated by local brick kilns. Since the vegetated top soil layer has removed from the ground, the exposed soils are appeared as bright spots in Landsat FCC image.

**Table 1** LULC in Delhi NCT and its Periurban area (2014)

| LULC classes | Delhi NCT (2014) | | Periurban area of Delhi NCT (2014) | |
|---|---|---|---|---|
| | Area in sq. km. | Percentage | Area in sq. km. | Percentage |
| High Density Built-up | 420.59 | 28.42 | 215.70 | 8.25 |
| Low Density Built-up | 281.38 | 19.01 | 558.16 | 21.34 |
| Dense Vegetation | 157.16 | 10.62 | 109.01 | 4.17 |
| Agricultural Land | 399.01 | 26.96 | 1199.26 | 45.84 |
| Sparse Vegetation | 124.51 | 8.41 | 287.92 | 11.01 |
| Waterbody | 21.79 | 1.47 | 21.73 | 0.83 |
| Wasteland | 75.62 | 5.11 | 224.28 | 8.57 |
| Total | 1480.07 | 100.00 | 2616.06 | 100.00 |

The areas under different land use and land cover classes were estimated and shown in table 1 and figure 3. The high density built-up area was found as dominant class in Delhi NCT region whereas agricultural land occupies the dominant position in the periurban area. It was estimated that about 420.59 sq. km. (28.42 %) of the total NCT region was under high density built-up land which, together with the low density built-up area (19.01 %) represents almost half of the total area. Whereas it depicts a different scenario in the periurban area where high density and low density built-up class covers an area of 215.70 (8.25 %) and 558.16 (21.34

%) sq. km. respectively. The estimated area under agriculture was about 399.01 (26.96 %) and 1199.26 (45.84 %) sq. km. in respectively for Delhi NCT and its periurban area. Although the percentage of sparse vegetation was comparatively higher in the periurban area (11.01 %) than the NCT region (8.41 %), the percentage of dense vegetation was found remarkably less in the periurban area (4.17 %) than the NCT region (10.62 %). This is due to absence of dense vegetation patches in periurban area. Unlike the periurban area, Delhi NCT region having many distinct patches of dense vegetation regardless of its high rate of urbanization.

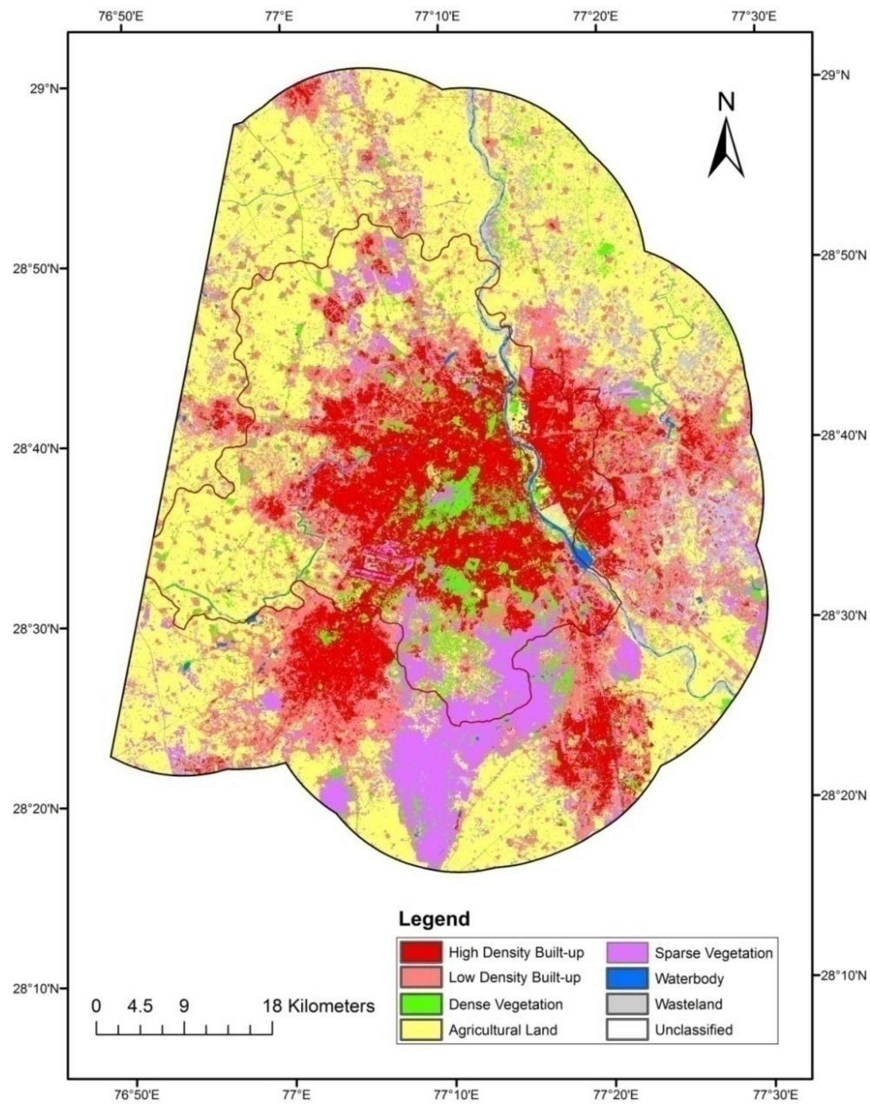


**Figure 2** Land use and land cover in 2014

**Table 2** Classification accuracy and Kappa statistics

| Land use and land cover Classes | 2014 | |
|---|---|---|
| | Producers Accuracy (%) | Users Accuracy |
| High Density Built-up | 97.92 | 94.00 |
| Low Density Built-up | 93.88 | 92.00 |
| Dense Vegetation | 84.91 | 90.00 |
| Agricultural Land | 97.87 | 92.00 |
| Sparse Vegetation | 88.46 | 92.00 |
| Waterbody | 97.96 | 96.00 |

| | | |
|---|---|---|
| Wasteland | 86.54 | 90.00 |
| Overall Classification | **0.92** | |
| Overall Kappa Statistics | **0.91** | |

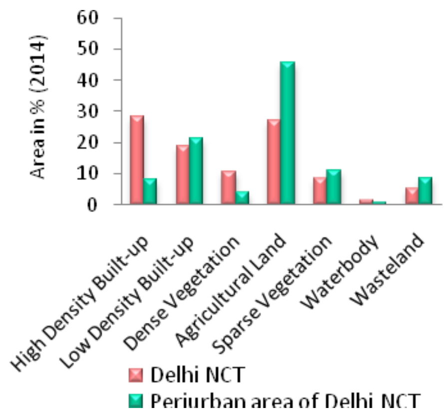


**Figure 3** Comparison of LULC in Delhi NCT vs. its Periurban area

### 3.2. Accuracy Assessment

In order to examine the performance of the classifier, the accuracy of the classified LULC map was assessed (table 2). It can be seen that the accuracy of each land use and land cover class of the classified LULC map was significantly high. It is also noteworthy that percentages of overall classification accuracy (0.92) as well as Kappa statistics (K^=0.91) were highly satisfactory. The accuracy of the SVM classifier with principal components reveals that there is a close agreement between classified and referenced data.

## 4. CONCLUSION

The major goal of the study was to assess the potentiality of Support Vector Machine (SVM) technique for identifying the spatial pattern of land use and land cover of Delhi NCT and its periurban area. The accuracy of the SVM classifier reveals that it is having enough potentiality to extract the accurate information of land use and land cover of an area. Acquiring accurate information of urban landscape using satellite images especially from a medium resolution image is a challenging task. From this point of view, SVM can be a better option for classifying medium resolution Landsat images. The study finds that the areas in Delhi NCT are highly urbanized but the urbanization in its periurban area has occurred around few centres like Gurgaon, Faridabad, Noida etc. The built-up areas in periurban zone are not uniformly distributed rather they are mainly concentrated in few locations. For mapping such heterogeneous pattern of land use and land cover, SVM proves to be a reliable technique.

## REFERENCE

1. Burges, C.J.C., (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery,* 2 (2), 121–167.
2. Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J. et al. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery, *Geoscience and Remote Sensing, IEEE Transactions on*, 42, 1530-1542.
3. Da Costa, S. M. F., Cintra J. P. (1999). Environmental Analysis of Metropolitan Areas in Brazil, *ISPRS Journal of Photogrammetry & Remote Sensing*, 54, 41-49.
4. Foody, G. M., Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines, *Geoscience and Remote Sensing, IEEE Transactions on*, 42, 1335-1343.
5. Griffiths, P., Hostert, P., Gruebner, O. et al. (2010). Mapping megacity growth with multi-sensor data, *Remote Sensing of Environment*, 114, 426-439.
6. Guo, Q., Kelly, M., Graham, C. H. (2005). Support vector machines for predicting distribution of Sudden Oak Death in California, *Ecological Modelling*, 182, 75-90.
7. Huang, C., Davis, L. S., and Townshed, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23, 725–749.
8. Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features", *Lecture Notes in Computer Science*, 1398, Springer-Verlag.
9. Kushwaha, S.P.S., Subramanian, S.K., Kameswara Rao, S.V.C.

et al. (1996). Interfacing remote sensing and GIS methods for sustainable development. *Int. J. Remote Sensing*, 17(15), 3055-3069.

10. Lu D, Weng Q. A. (2007). Survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*. 28, 823–870
11. Mahesh, P., Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86, 554–565.
12. Melgani, F., Bruzzone, L. (2004). Classification of Hyperspectral Remote Sensing Images With Support Vector Machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1778-1790.
13. Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 247-259.
14. Nooni, I. K., Duker, A. A., Van Duren, I. et al. (2014). Support vector machine to map oil palm in a heterogeneous environment, *International Journal of Remote Sensing*, 35, 4778-4794.
15. Pal, M., Mather, P. M. 2005. Support vector machines for classification in remote sensing, *International Journal of Remote Sensing*, 26, 1007-1011.
16. Poursanidis, D., Chrysoulakis, N., Mitraka, Z. 2015. Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping, *International Journal of Applied Earth Observation and Geoinformation*, 35, 259-269.
17. Vapnik. V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.