

RESEARCH ARTICLE



GOPEN ACCESS

Received: 15-04-2020 Accepted: 09-04-2020 Published: 07-06-2020

Editor: Dr. Natarajan Gajendran

Citation: R V, A M, S V, R P (2020) Unsupervised ISODATA algorithm classification used in the landsat image for predicting the expansion of Salem urban, Tamil Nadu. Indian Journal of Science and Technology 13(16): 1619-1629. https://doi.org/ 10.17485/IJST/v13i16.271

***Corresponding author**. Vimala R

Department of Computer Science, Government Arts College for Women, Salem, India vimalagovtarts@gmail.com

Funding: University Grants Commission (UGC), Government of India

Competing Interests: None

Copyright: © 2020 R, A, S, R. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

Unsupervised ISODATA algorithm classification used in the landsat image for predicting the expansion of Salem urban, Tamil Nadu

Vimala R^{1*}, Marimuthu A², Venkateswaran S³, Poongodi R³

Department of Computer Science, Government Arts College for Women, Salem, India
 Department of Computer Science, Government Arts College, Coimbatore, India
 Department of Geology, Periyar University, Salem, India

Abstract

Objectives: To study the land cover change Salem city as a case study of urban expansion in India covering the span of 35 years from 1990 to 2025. **Method:** Remote sensing methodology is adopted to study the geographical land use changes occurred during the study period (year 1990- 2025). Landsat images of TM and ETM+ of Salem city area are collected from the USGS Earth Explorer website. After image pre-processing, unsupervised image classification has been performed to classify the images into different land use categories. Seven land use classes have been identified as road, urban (Build up), vegetation, water bodies, fallow land, mines and barren land. Classification accuracy is also estimated using the field knowledge obtained from field surveys. Findings: The obtained accuracy is between 83 to 86% of all the classes. Change detection analysis shows the built-up area has been increased by 1.49 km², vegetation area has been decreased by 11.55km². Application: Information on Urban growth, land use and land cover change study is very useful to local Government and Urban Planners for the betterment of the future plans of sustainable development of the city.

Keywords: Land use /Land cover; Urban Sprawl; Urbanization; Remote Sensing; Landsat data; Salem city

1 Introduction

A huge amount of remote sensing data by a traditional manual method is daunting. The advent of computers eliminates the manual data processing job by automatically processing and analyzing the remotely sensed data with the aid of powerful digital image processing software. Image classification is the process of grouping and labelling each pixel within the original image to a land use/cover information class^(1,2). Computer aided classification based on the pixel values are performed on the assumption that each spectral class corresponds to a spectral cover. This classification approach helps us to more quickly study the earth surface features from the image data and take necessary action immediately^(3,4). The Anderson's image classification scheme has been adopted

to list all the land use and land cover types within the study area using automated unsupervised classification techniques. Several researchers made an attempt to better understand the urban growth rate using remotely sensed data^(5,6). They found that built up area has sharply increased due to the construction of new buildings on agricultural and vegetation lands.

Urban expansion is the greatest challenge of recent century in the developing countries due to rapid population growth, economic development and infra structural development initiatives. Urban sprawl has an important area of research all over the world have described the urban sprawl development⁽⁷⁾. This phenomenon usually takes place either in a radial direction from the center of the city or in a linear direction along the highways and drives the change in land use patterns. Urban sprawl has been analyzed for the inefficient use of land resources and energy and large-scale encroachment on agricultural land. Several researches have evaluated in this case for the loss of open space, environmental damage, loss of surface water bodies, and depletion in groundwater level. The unplanned urban growth and expansion had serious impacts on the urban ecosystem and on the sustenance of natural resources. Rapid urbanization with high population density often faces a severe crisis due to inadequate infrastructure and lack of basic amenities. In order to achieve proper management of urban sprawling, the measurement, mapping and monitoring of urban sprawl are crucial for the government officials and planners in any region⁽⁸⁾. The reliable and updated information on spatio-temporal pattern of urban sprawl is a prerequisite for the sustainable urban development planning and management. The built up is the parameter for quantifying urban sprawl. The extract information on spatialtemporal changes of the Salem urban helps urban managers to focus their attention on the areas that are under expansion or to properly manage natural resources and monitor local environmental changes associated with the urban sprawl. It is achieved by analyzing LANDSAT images for the years of 1990, 1995, 2000, 2005, 2010, and 2015 using unsupervised ISODATA algorithm classification techniques.

2 Profile of the Salem Urban

Salem urban is one of the rapidly growing city in Tamil Nadu for the past three decades owing to anthropogenic influences. The city is bounded by hill and hillocks, the Shevaroy hill ranges in the north, the Kanjamalai hill in the southwest, Chalk hills in the west, the Nagaramalai in the North-West, the Kumaragiri hill in the southeast. It falls within the Geo-coordinates ranging between 11°35' 50" N to 11°42'10" N and 78°5'0" E to 78° 14'30" E. The urban area comprises of 60 wards constituting 4 Zones named as Suramangalam, Hasthampatty, Ammapet and Kondalampatty Zones with an aerial extent of 338.38 Km². It experiences temperature ranging from 20°C to 37.9°C and the temperature is usually very high during summer. It receives 363.5mm rainfall as an average annual rainfall⁽⁹⁾. From the 2001 census report, the total population of the Salem urban is about 30,16,346 of which 12,79,846 are employed and the rest are unemployed. Railway junction is situated at the center of the urban. The interior of the urban area is well-connected by road network and has well-established connections with the adjacent cities including Bangalore, Chennai, Trichy and Coimbatore through the rail and highway road networks of NH47, and NH38. The location map of Salem urban is shown in Figure 1.



Fig 1. Location map of the Salem urban

3 Unsupervised Classification

Unsupervised classification is useful for scenes in which land cover is not well-known or undefined. This classification is supposed to automatically categorize all pixels in an input image into land cover classes. Computer algorithms group similar pixels into various spectral classes which the analyst must then identify and combine into information classes^(10,11). In the unsupervised classification, also known as clustering analysis, pixels are aggregated into certain categories based on the similarity in their spectral values. In this type of classification, the image analyst does not need to know about the land covers in the study area before the clustering analysis as the algorithm of this classification reads pixels and aggregates them into a number of clusters, known as spectral classes. All the closely ranged DN values in each cluster, refer to a particular land cover. Then the land cover class is identified by comparing the classified spectral classes with the corresponding reference data. But in supervised classification, an analyst uses previously acquired knowledge of an area, or a priori knowledge, to locate specific areas, or training sites, which represent homogeneous samples of known land use and/or land cover types^(12,13). Based on statistics of these training sites, each pixel in an image is then assigned to a user-defined land use type (residential, industrial, agriculture, etc.) or land cover type (forest, grassland, paved surface, etc.). Among the two primary classification methods, unsupervised classification has been chosen in this research and the ISODATA classifier, one of the most popular classifiers in the unsupervised classification has been discussed below.

4 ISODATA Algorithm

The term ISODATA stands for Iterative Self-Organizing Data Analysis. It is a most common used method in unsupervised classification. The ISODATA is executed by using the following algorithm:

Step 1:Specify the total number of spectral classes to be clustered

Step 2: Arbitrarily select the total number of cluster centers as the candidates

Step 3: Calculate the distance of every pixel in the input image to each of the cluster centers using the Euclidean spectral distance method, mathematically expressed as below in equation (1):

$$De = \sqrt{\sum_{i=1}^{n} \left(DN_{Bi} + DN_{Ai} \right)^2} \tag{1}$$

Where, n = number of spectral bands used in a classification

 DN_{Ai} = DN of pixel A in the ith band

 DN_{Bi} = DN of pixel B in the same band, i.e., band i

Step 4: Identify pixels which are closer to each cluster centre

Step 5: Assign all of those pixels to the corresponding cluster centres

Step 6: Calculate the Sum of Squared Error(SSE) from the pixels of the respective clusters using the following equation (2):

$$SSE = \sum_{j=1}^{k} \sum_{i=1}^{n} [DN(i,j) - m_j]^2$$
(2)

Where n = number of pixels enclosed in a given cluster. Its specific value varies from cluster to cluster, DN (i, j) = value of the ith pixel in the jth cluster, m_j = mean of the jth cluster

Step 7: Adjust the centre of each cluster if SSE is high and update it until the SSE reaches the specified minimum value Step 8: Output the image classification results for post classification analysis

Algorithm: ISODATA

The analyst tells the algorithm, how many clusters should be formed using the input image data. The algorithm mathematically locates the specified number of cluster centres as the candidates. The distance between every pixel and the cluster centers is measured by the Euclidean spectral distance method. The pixels which are in shortest spectral distance to the cluster centers are assigned to the cluster. Similarly, all the pixels in the image are assigned to all the corresponding clusters. Following the pixel assignment to the clusters, the Sum of Squared Error (SSE) is found from the pixels of the respective clusters. If SSE is high, the center of each cluster is adjusted and updated through the process of iteration until the SSE reaches the specified minimum value⁽¹⁴⁾. One of the main advantages of this ISODATA algorithm is that it automatically deletes, merges, and splits using a mathematical function. The algorithm deletes a cluster/clusters, when the number of pixels in clusters is less than the specified threshold values. It merges two clusters if spectral distance between them is shorter than the predetermined threshold and the algorithm splits the early formed cluster if it contains a large number of pixels exceeding the specific limit.

5 Methodology

Unsupervised classification has been carried out using the ISODATA algorithm. According to the methodology given in Figure 2, the first four principal component images were fed into the unsupervised classifier. The number of clusters to be classified was set at 60 and the maximum number of iterations was ten iterations. The convergence threshold to be assigned as 0.95 to stop the algorithm as soon as 95% or more of the pixels stayed in the same cluster between two iterations. After the completion of the clustering process, the program defined 60 clusters were reduced into 7 clusters corresponding to the number of LU/LC classes with the aid of field knowledge and supporting data including Google Earth images, and PCA-FCC images. Finally, the LU/LC thematic maps were produced as outputs by the unsupervised classifier.



Fig 2. Flowchart for unsupervised classification

6 Landsat Images and their Characteristics

Six Landsat images are shown in Figure 3 were downloaded from https://landsat.usgs.gov/website to study the LU/LC change of the study area. Out of the six images, the four images belonging to the years of 1990, 1995, 2000, and 2010 belong to Thematic Mapper (TM) sensor. The other two input images of 2005 and 2015 are from Enhanced Thematic Mapper plus (ETM+) sensor and Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) respectively. The characteristics of the Landsat images of different sensors (USGS Website) including number of bands, their wavelength, and spatial resolution have been shown in the Tables 1, 2 and 3. Spectral bands of the six images were stacked together to form a single composite image file. Then the output images were used for PCA in order to choose the most informative bands regarding the scene.

Table 1. Characteristics of Landsat E 1 M+ Image					
Type of Images	Colors Bands	Wavelength in micrometers	Resolutionin meters		
	Band 1 – Blue	0.45-0.52	30		
	Band 2 - Green	0.52-0.60	30		
	Band 3 – Red	0.63-0.69	30		
Landsat 7 EnhancedThematic MapperPlus(ETM+)	Band 4 - Near Infrared (NIR)	0.77-0.90	30		
	Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30		
	Band 6 - Thermal	10.40-12.50	60 * (30)		
	Band 7 - Shortwave Infrared (SWIR) 2	2.09-2.35	30		
	Band 8 - Panchromatic	.5290	15		

* ETM+ Band 6 is acquired at 60-meter resolution, but products are resampled to 30-meter pixels.

Type of Images	Bands	Wavelengthin micrometers	Resolutionin meters
	Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
	Band 2 – Blue	0.452 - 0.512	30
	Band 3 – Green	0.533 - 0.590	30
	Band 4 – Red	0.636 - 0.673	30
Landsat 8	Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
OperationalLand Imager(OLI) and Thermal	Band 6 - Shortwave Infrared (SWIR) 1	1.566 - 1.651	30
	Band 7 - Shortwave Infrared (SWIR) 2	2.107 - 2.294	30
	Band 8 - Panchromatic	0.503 - 0.676	15
	Band 9 – Cirrus	1.363 - 1.384	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

Table 2. Characteristics of Landsat 8 Image of OLI and TIRS

* TIRSbands are acquired at 100 meter resolution, but are resampled to 30 meter indelivering data product.



Landsat 8 image in the year 2015

Fig 3. Landsat TM and ETM+ images of Salem Urban from the year 1990 to 2015

Table 3. Characteristics of Landsat TM Image					
Type of Images	Bands	Wavelengthin micrometers	Resolutionin meters		
Landsat 4-5 Thematic Mapper(TM)	Band 1 – Blue	0.45-0.52	30		
	Band 2 – Green	0.52-0.60	30		
	Band 3 – Red	0.63-0.69	30		
	Band 4 - Near Infrared (NIR)	0.76-0.90	30		
	Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30		
	Band 6 – Thermal	10.40-12.50	120* (30)		
	Band 7 - Shortwave Infrared (SWIR) 2	2.08-2.35	30		

* TM Band 6 was acquired at 120-meter resolution, but products are resampled to 30-meter pixels.

7 Supporting Data

The unsupervised classifier produced 60 clusters using the Landsat images as inputs. As this research required only 7 LU/LC classes, the 60 clusters were then reduced into 7 LU/LC classes using the supporting images of the corresponding year. The supporting images, including PCA-FCCs and their equivalent Google Earth images used during the unsupervised image classification process, they have been shown in Figure 4. These image sets have also been used in preparing training data sets in order to train and test the Multilayer perceptron neural networks before classifying the input images. Google Earth images of the years 1990, 1995, 2000, 2005, 2010, and 2015 have also been used as the base maps for collecting ground truth to perform classification error evaluation which is the last step of image classification.



FCC-PCA image in the year 1990



Google Earth image in the year 1990.



FCC-PCA image in the year 1995



Google Earth image in the year 1995.



FCC-PCA image in the year 2000



Google Earth image in the year 2000



FCC-PCA image in the year 2005

Google Earth image in the year 2010



Google Earth image in the year 2005



ECC-PCA image in the year 2015



FCC-PCA image in the year 2010



Google Earth image in the year 2015

Fig 4. Falsecolor composite of principal components and Google earth images from the year1990 to 2015

8 Result and Discussion of Unsupervised Classification

The first four principal component images were used as inputs into the unsupervised classifier using the ISODATA algorithm. The classified images in Figure 5 produced by the classifier and the corresponding Table 4 showing the spatial extent of each land use and land cover have been discussed in the following paragraph.

In the year 1990, it was known that all of the urban areas accounted for only 23.64 km² (6.99%) of the total study area. Particularly, areas in and around Shevapettai and the areas at the road margins and at the junction point of the main roads have been heavily converted into urban areas. The spatial extent of vegetation and fallow land covers is about 137.07 km² (40.51%) and 106.03km² (31.33%) respectively. During 1990, the road network covered 2.39 km² (0.71%) and water bodies 1.61 km² (0.48%). Mining areas account for 11.81km² (3.49%) and barren land covers about 55.83km² (16.50%). In the year 1995, the road network occupied about 2.49km² (0.74%) and urban 24.1km² (7.12%). Urban areas have mostly developed on fallow lands. The vegetation covers a larger portion of the study area with the spatial extent of 149.93km² (44.31%), followed by fallow land with the spatial extent of 82.70 km² (24.44%). Water bodies cover 1.40 km² (0.41%), mines 12.50 km² (3.69%), and barren land 65.26 km² (19.29%).

In the year 2000 the road network occupied 2.50 km² (0.74%). The urban areas covered 24.27 km² (7.17%). The land cover that occupies the major part of the study area is vegetation showing the spatial extent of 191.1km² (56.47%). Water bodies reserved 2.71 km² (0.80%), fallow land 65.41 km² (19.33%), mines 13.12 km² (3.88%), and barren land 39.27 km² (11.61%). It can be noted that during 2000, new water bodies have emerged in several places and vegetative cover has occupied more land space when compared to the year 1990. In the year 2005, it is found that the road cover occupies about 3.30 km² (0.98%). Urban areas reserved 38.38 km² (11.34%), vegetation 73.33 km² (21.67%), water bodies 3.98 km² (1.18%), fallow land 149.53 km² (44.19%), mines 11.19 km² (3.31%), and barren land 58.67 km² (17.34%). Most importantly, urban areas have developed on the lands at the margin of the already developed urban areas and on the lands behind the developed areas along the road margins.

In the year 2010, the spatial extent of each LU/LC indicated that road occupied 3.69km² (1.09%), urban 56.27km² (16.63%), vegetation 87.66km² (25.91), water bodies 1.9km² (0.56%), fallow land 106.35km² (31.43%), mines 14.89, and barren land 67.62km² (19.98%). In the year 2015 that road had covered 3.70 km², urban 57.75 km², vegetation 95.51 km², water bodies 5.45 km², fallow land 98.28 km², mines 14.75 km², and barren land 62.94 km². From the observation of the classified LU/LC image of 2015, it is found that mining areas have been developed in the extreme north of the study area. A water body near Kannankurichi have been shrunken by the urban development activities. The urban areas have been expanding more and more from the center of the already developed urban areas.







LU/LC image in the year 2000



LU/LC image in the year 2010



LU/LC image in the year 1995



LU/LC image in the year 2005



LU/LC image in the year 2015

Fig 5. LU/LC images of unsupervised classifier from the year 1990 to $2015\,$

LU/LC Class	In the 1990LU/LC Km ²	In the 1995LU/LC Km ²	In the 2000LU/LC Km ²	In the 2005LU/LC Km ²	In the 2010LU/LC Km ²	In the 2015LU/LC Km ²	In the 2025LU/LC Km ²	Predicted change between 2015 and 2025 km ²
Road	2.39	2.49	2.50	3.30	3.69	3.70	5.19	▲1.49
Urban	23.64	24.1	24.27	38.38	56.27	57.75	66.47	▲8.72
Vegetation	137.07	149.93	191.1	73.33	87.66	95.51	83.96	▼11.55
Water Bod- ies	1.61	1.40	2.71	3.98	1.9	5.45	6.03	▲0.58
Fallow Land	106.03	82.7	65.41	149.53	106.35	98.28	94.08	▼4.2
Mines	11.81	12.5	13.12	11.19	14.89	14.75	14.71	▼0.04
Barren Land	55.83	65.26	39.27	58.67	67.62	62.94	62.94	0
Total	338.38	338.38	338.38	338.38	338.38	338.38	338.38	

Table 4. Unsupervised classification's land use/ Land cover status of Salem urban in the years between 1990 and 2025.

Note: \blacktriangle denotes positive LU/LC change and \blacktriangledown denotes negative LU/LC change

9 Conclusion

The simulated predicted model results for the year 2025 are given in Fig. 5. It reveals that urban areas will have moved up from 57.75 to 66.47 km² with an increase of about 8.72 km². Road network will have increased by about 1.49 km² and water bodies 0.58 km², while vegetation and fallow land will have decreased by about 11. 55 and 4.2 km² respectively.



Fig 6. LU/LC change simulated model for the year2025

It is apparent that urban expansion will occur at the expense of vegetation and fallow land. Moreover, the model predicts that mines will have decreased from 14.75 km² to 14.71 km² by the year of 2025. Barren land shows zero variation because it is taken to be constant. The fig. 6 shows that urban expansion will prominently occur in areas adjacent to the already developed areas and in areas adjacent to the road network.

Salem urban is naturally bounded by hills, hillocks and ridges which plays an important role in controlling the city expansion. Normally the city expansion is controlled by several factors such as employment, educational facilities, healthcare facilities, road networks and administrative support. In addition to this, the supplementary factors include land topography, availability of groundwater and its quantity and quality, nature of soil and basement rock, environmental issues, community settlement, real estate promoters etc. Since the urban expansion is duly controlled by Shevaroy hill, Kumaragiri hill, Chalk hill and Nagaramalai hill, the liner probability of Salem urban expansion will be towards the North-East and South-West directions.

Acknowledgment

The first author feels grateful to the University Grants Commission (UGC), Government of India, for the financial and infrastructure support; further the authors are thankful to the United States of Geological Survey-USGS, Earth Explorer website for the satellite data source and Salem City Corporation for providing necessary data for this research work.

References

- 1) Eastman J, Fulk M. Long Sequence Time Series Evaluation Using Standardized Principal Components. . *Photogrammetric Engineering & Remote Sensing*. 1993;59(8):1307–1312. Available from: http://worldcat.org/issn/00991112.
- 2) Faridkhan T, Vijayalakshmi T, Kumar S. Land use land cover change detection of Ghatkesarmandal, Ranga Reddy district using Remote sensing and GIS. International Journal of Scientific & Engineering Research. 2012;3(11). Available from: https://www.ijser.org/researchpaper/Land-use-land-cover-change-detection-of-Ghatkesar-mandalRangaReddy-district-using-Remote-sensing-and-GIS.pdf.
- 3) Shankar B. Chromium Pollution in the Groundwaters of an Industrial Area in Bangalore, India. *Electronic Green Journal*. 2009;26:305-310. doi:10.1089/ees.2008.0043.
- 4) Shalaby A, Tateishi R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography*. 2007;27(1):28–41. doi:10.1016/j.apgeog.2006.09.004.
- 5) Vital JA. Land Use/Cover change using Remote Sensing and Geographic Information Systems: Pic Macaya National Park, Haiti. Pic Macaya National Park, Haiti. 2008.
- 6) Rawat JS, Kumar M. Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*. 2015;18(1):77–84. doi:10.1016/j.ejrs.2015.02.002.
- 7) Ismail MH, Jusoff K. Satellite Data Classification Accuracy Assessment Based from Reference Dataset. *World Academy of Science, Engineering and Technology*. 2008;2(3):527–533. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.307.4674&rep=rep1&type=pdf.
- 8) Song C, Woodcock CE, Seto KC, Lenney MP, Macomber SA. Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects. *Remote Sensing of Environment*. 2001;75(2):230–244. doi:10.1016/s0034-4257(00)00169-3.
- 9) Vimala R, Marimuthu A. Spatial Temporal Expansion of Urban Sprawl in the Salem City of Tamil Nadu, India Using Geospatial Techniques. *International Journal of Computer Application*. 2016;6(2):22–27. Available from: https://rspublication.com/ijca/2016/april16/4.pdf.
- 10) Thomson AG, Fuller RM, Eastwood JA. Supervised versus unsupervised methods for classification of coasts and river corridors from airborne remote sensing. *International Journal of Remote Sensing*. 1998;19(17):3423–3431. doi:10.1080/014311698214091.
- 11) Jensen JR. A Remote Sensing Perspective. In: Introductory Digital Image Processing. Upper Saddle River, NJ. Pearson Prentice Hall. 2005.
- 12) Sundar ML, Prasath CA, Rosario HE, Tamilselvan K. Urban Sprawl Mapping and Land Use Change Detection Analysis. *International Journal on Advanced Science, Engineering and Information Technology*. 2019;9(1):281–281. doi:10.18517/ijaseit.9.1.8080.
- 13) Shrivastava RJ, Gebelein JL. Land cover classification and economic assessment of citrus groves using remote sensing. *ISPRS Journal of Photogrammetry* and Remote Sensing. 2007;61:341–353. doi:10.1016/j.isprsjprs.2006.10.003.
- Mandal J, Ghosh N, Mukhopadhyay A. Urban Growth Dynamics and Changing Land-Use Land-Cover of Megacity Kolkata and Its Environs. *Journal of the Indian Society of Remote Sensing*, 2019;47(10):1707–1725. doi:10.1007/s12524-019-01020-7.