# Application of Classification Techniques for Identification of Water Region in Multiple Sources using Landsat-8 OLI Imagery

Vikash K. Mishra[1], and T. Pant[2]

(1) Indian Institute of Information Technology, Allahabad-211015, U.P., India.
(2) Indian Institute of Information Technology, Allahabad-211015, U.P., India.

## Abstract

Landsat-8 OLI (Operational Land Imagery) data is used to identify the water of the Ganges and the Yamuna in the Sangam region of Allahabad of Uttar Pradesh, India. Landsat-8 has eleven different bands and each band has its own significance in terms of remotely sensed information. The study exhibits the suitable band from these available eleven bands to classify the water of two rivers in the same class more accurately, either we use supervised or unsupervised classification schemes. The study will help to choose that band which will identify the water in such regions where different sources of water used to meet.

## 1. Introduction

Sangam is a place where two great rivers of India the Ganges and the Yamuna used to meet. The two rivers maintain their visible identity and can be identified by their different colors. The water of the Ganges is clear, while that of the Yamuna is greenish in color [1].

Satellite images are rich enough to provide the information of geographical area, which provide both quantitative and qualitative information that reduces human efforts in terms of field survey. Image classification is a process of grouping similar pixels into meaningful classes. Satellite image classification can also be thought as extracting information from satellite images. There is a strong need of effective and efficient mechanisms to extract and interpret valuable information from massive satellite images.

Remote sensing is a science and art of getting information about some object, method or phenomenon through the analysis of data acquired by a device without any contact [2]. Remote sensing can be thought of as a reading process in which eyes work as a sensor to sense the electromagnetic energy reflected by text and collect it as data which is further processed by mind to form the words. Sun is the best source of electromagnetic energy. When this energy falls on various earth surface features, actually nine in numbers: Urban or Built-up Land, Agricultural Land, Rangeland Forest Land, Water, Wetland, Barren Land, Perennial Snow or Ice and Tundra also termed as Land use Land Cover (LULC) [3], reflects and emits electromagnetic energy then either airborne or space-borne electromagnetic sensor acquire data either in pictorial or digital form. Electromagnetic remote sensing

of earth surfaces involved two basic steps data acquisition and analysis [2].

### 1.1 Data Acquisition

In order to acquire the data remotely following elements are required: energy source, propagation of energy through atmosphere, energy interaction with earth surface features, retransmission of energy through atmosphere, airborne/space-borne sensor and generation of sensor data in pictorial or digital form. Figure 1 shows the acquisition process.

### 1.2 Analysis of acquired data

The analysis of acquired data through sensor requires following: examining the data using various viewing and interpretation devices information is extracted, information is compiled in the form of maps, table and computer file etc. and finally information is presented to user, who apply it to their decision making process.
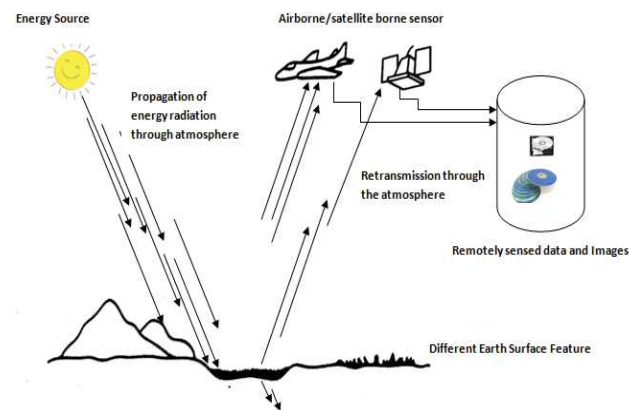


Figure 1.Remote Sensing of Earth Surface Features

The Landsat era began in 1972, from the United States Geological Survey (USGS) is one of the best sources for mapping and monitoring of land cover and land surface biophysical and geophysical properties over the last 40 years. Landsat-8 was successfully launched on February 11th 2013, which is the latest satellite in the Landsat series offers scientists a clearer view with better spatial resolution than most ocean-sensing instruments and greater sensitivity to brightness and color than previous Landsats [4]. Landsat-8 carries two instruments: The Operational Land Imager (OLI) (from band1 to band9), excluding band9, is used to study the different earth

surface features which are Nine in numbers according to USGS[3] and band9 is used for Improved detection of cirrus cloud contamination [4]. Other is Thermal Infrared Sensor (TIRS) including band10 and band11 [4][5]. The difference between water of Ganges and Yamuna can also be seen in the classification of the landsat-8 imagery of Sangam Region, so we have included only band1 to band8 [4] to identify the water in same class as only these bands are capable to identify LULC [3] of the earth surface.

## 2. Study Area and Data Used

Allahabad is a small district in the state of Uttar Pradesh, India. The Triveni Sangam or Sangam is the place where two rivers Ganges and Yamuna used to meet. Landsat-8 imagery from the United States Geological Survey (USGS) is downloaded free from https://libra.developmentseed.org/ to continue the study. We have taken April, 2013 and May, 2018 data set from the above mentioned source and further resized to a new coordinate as: for April, 2013 upper left corner, latitude is $25^0 28' 18.37"$N, longitude is $81^0 48' 33.54"$E and lower right corner, latitude is $25^0 23' 5.67"$N, longitude is $81^0 58' 13.82"$E. Figure 2 shows the data set of April, 2013.



(a)     (b)     (c)
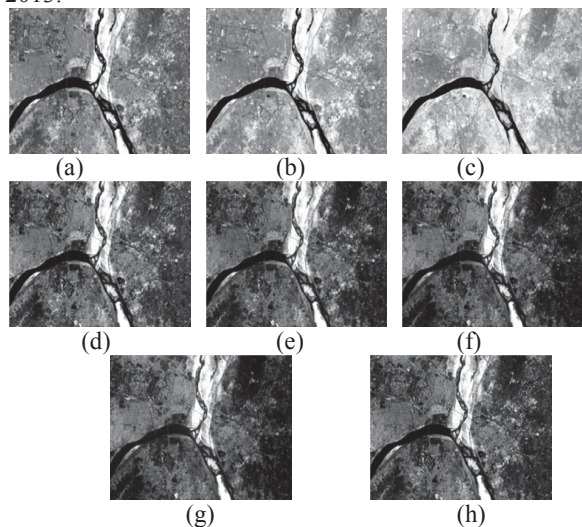
(d)     (e)     (f)

(g)     (h)

Figure 2.Operational Land Imagery (band 1 to 8) of Landsat-8 of May 2018 data set

For May, 2018, upper left corner, latitude is $25^0 27' 19.84"$N, longitude is $81^0 51' 28.24"$E and lower right corner, latitude is $25^0 23' 53.69"$N, longitude is $81^0 55' 12.25"$E. Figure 3 shows the data set of May, 2018.



(a)         (b)         (c)         (d)
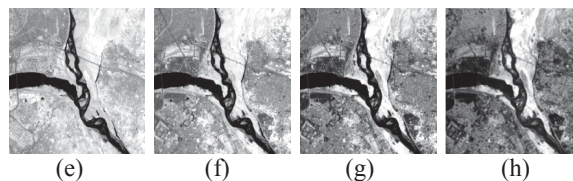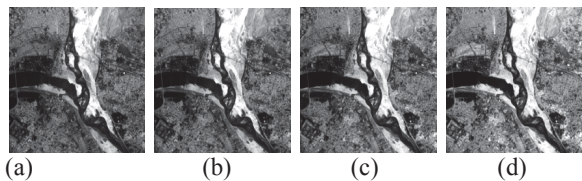


(e)         (f)         (g)         (h)

Figure 3.Operational Land Imagery (band 1 to 8) of Landsat-8 of May 2018 data set

## 3. Methodology

For the identification of water in Sangam region over different bands of landsat-8 imagery we used both the schemes i.e., Supervised Classification and Unsupervised Classification [2].

### 3.1 The Supervised Classification
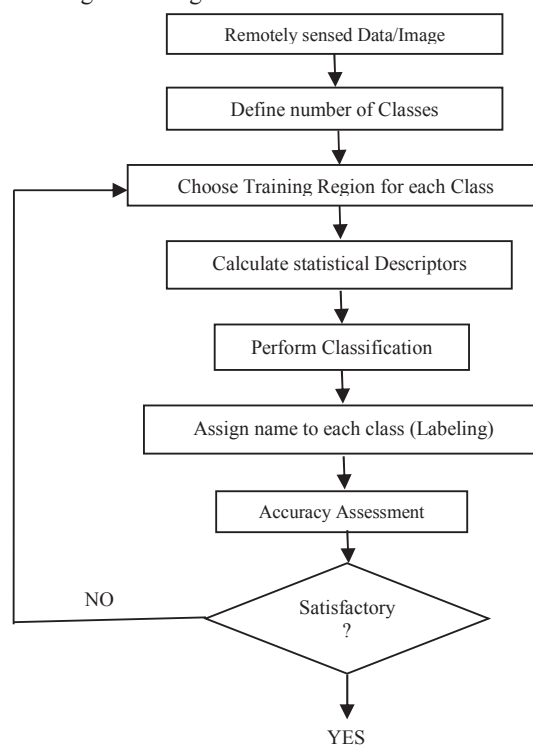
This scheme is given in Fig.4.



Figure 4.Supervised Classification Scheme

Minimum Distance [6][8] supervise classification scheme is used, where the average spectral value in each band for each class is computed. The spectral value of the pixel is compared to the means of each category and assigned to the closest class. Following steps are recommended to implement the above defined scheme:

1. Refer the Remotely sensed data or image.
2. Define the possible number of classes and select Training Data Set for each Class.
3. Calculate min value for each class on the basis of Training data set.
4. Take a unclassified pixel find its distance to mean value of different classes and classify into a class having minimum distance to mean.

5. Repeat Step 3 for all unclassified pixels and classify them into appropriate Class.
6. After classification give name to each class.

## 3.2 The Unsupervised Classification

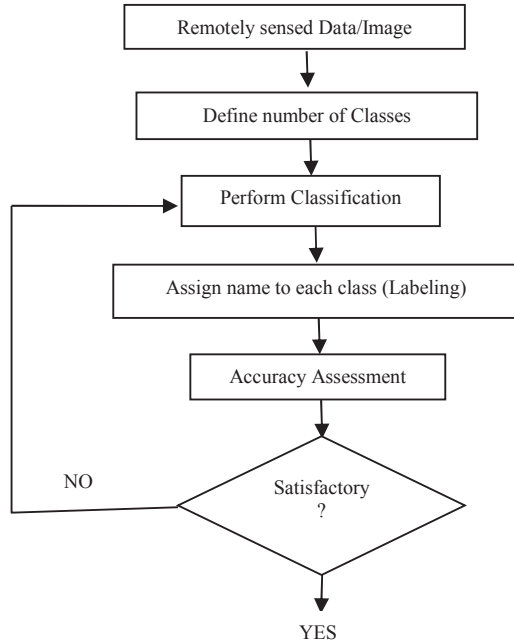This is a pixel based and computer automated classification approach. This Approach is depicted in Figure 5.



Figure 5.Unsupervised Classification Scheme

ISODATA [6][7][8] unsupervised scheme is used. This method doesn't require the prior information about number of clusters. This Algorithm splits and merges classes on the basis user defined threshold values. This Algorithm runs iteratively to satisfy the threshold value. The Algorithm is as follows:

1. Classes centres i.e. mean are placed randomly.
2. Pixels are assigned to classes based on the minimum distance to mean method
3. The standard deviation within each cluster and distance between two centres are computed.
   3a. classes are split if the standard deviation is greater than the user defined threshold value.
   3b. Classes are merged together if the distance between centres is less than user threshold values.
4. New iteration is performed with new class centre.
5. Iterations are performed until the average inter centre distance [2][7] falls below the user defined threshold or the maximum number of iteration is reached.

## 3.3 Inspection

After Classification, classified image is inspected on the basis of Visual Attributes and Confusion Matrix. This will be done through following steps.

1. Select the one band from landsat-8 data set and classify it, say it as classified band.
2. In the classified band, if water of Ganges and Yamuna are in same class then go to step-3, else discard the current band and choose the next band from the dataset and repeat the steps form beginning.
3. Draw confusion Matrix [2][7][8] for the classified band and find out the overall accuracy.
4. Repeat step 1 to 3 for all bands having water of Ganges and Yamuna in same class and compare the overall accuracy.

## 4. Result and discussion

We have used "Environment for Visualizing Images" (ENVI 5.1) software tool to classify the Landsat-8 Imagery from April, 2013 and May, 2018. We considered each band from eight different bands one by one to identify the water of two rivers Ganges and Yamuna. Minimum Distance Supervised Classification applied on Landset-8 data set and then inspection is done, bands 1, 2, 3, 4, and 8 are not identifying the water of Ganges and Yamuna as same class shown in fig. 6, while band 5, 6 and 7 are identifying the water of two rivers in same class as it is reflected in figure 7. Table 1 is showing the classification accuracy [2][7] regarding each band.
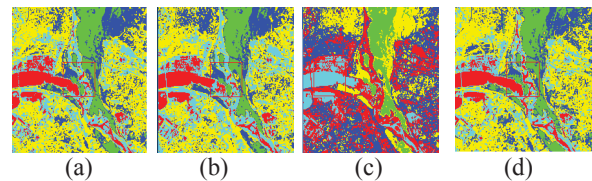


Figure 6.Classified bands 1, 2, 3 and 4 in a, b, c and d using Minimum Distance Supervised Classification of May 2018 data set
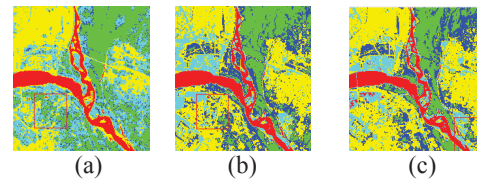


Figure 7.Classified bands 5, 6 and 7 in a, b and c Minimum Distance Supervised Classification of May 2018 data set

Table 1
Accuracy assessment for Minimum Distance Supervised Classification for band5, band6 and band7

| Band | April, 2013 data set | | May, 2018 data set | |
|---|---|---|---|---|
| | Overall accuracy (%) | Accuracy of Water class (%) | Overall accuracy (%) | Accuracy of Water class (%) |
| Band5 | 44.20 | 97.50 | 50.91 | 95.87 |
| Band6 | 64.37 | 96.50 | 72.32 | 95.41 |
| Band7 | 72.62 | 95.75 | 81.38 | 94.04 |

ISODATA Unsupervised classification was applied and then inspection is done, bands 1, 2, 3, 4 and 8 are not identifying the water of Ganges and Yamuna as same class shown in Figure 8., while band 5, 6 and 7 are identifying the water of two rivers in same class shown in figure 9., Table 2 is giving the overall accuracy and individual accuracy of each class.
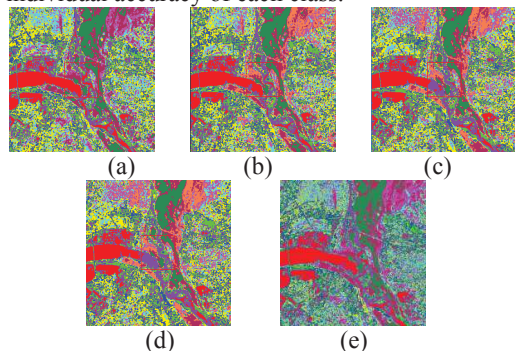


(a)      (b)      (c)

(d)      (e)

Figure 8.Classified bands 1, 2, 3, 4 and 8 in a, b, c, d and e using ISODATA unsupervised Classification of May 2018 data set.
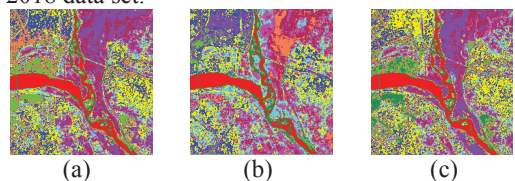


(a)      (b)      (c)

Figure 9.Classified bands 5, 6 and 7 in a, b and c using ISODATA unsupervised Classification of May 2018 data set

It can be seen from figure 7 and figure 9 which are actually the classified images of band 5, 6 and 7 where water is classified and identified in red color while in figure 6 and figure 8 classified images of band 1, 2, 3, 4 and 8 are not able to identify water bodies available in top and bottom region. Similar visual results are also obtained for data set of April, 2013 and only band 5, 6 and 7 are identifying the water and from table 1 and table 2 the accuracy can also be seen.

Table 2
Accuracy assessment for ISODATA Unsupervised Classification for band5, band6 and band7

| Band | April, 2013 data set | | May, 2018 data set | |
|------|-------------------------|-------------------------|-------------------------|-------------------------|
| | Overall accuracy (%) | Accuracy of Water class (%) | Overall accuracy (%) | Accuracy of Water class (%) |
| Band5 | 33.36 | 97.00 | 44.63 | 94.97 |
| Band6 | 32.49 | 97.59 | 41.20 | 95.72 |
| Band7 | 34.72 | 97.92 | 56.00 | 98.98 |

**5. CONCLUSION**

In this study, two data sets from Landsat- 8 are taken, one is from April 2013 and other one is from May, 2018. We applied supervised and unsupervised classification on each band (Operational Land Imagery) i.e., from band 1 to band 8. On classified data visual and statistical both kind of inspection is performed. In visual inspection, classified band is analyzed with perspective that water of both the rivers must classify in same class i.e., same color is assigned to water bodies. If it is so we go for statistical analysis. In statistical analysis accuracy assessment is done on the basis of confusion matrix. Band 1, 2, 3, 4 and 8 are classified and water is obtained as different class for both the rivers, while Band 5, 6, and 7 are classifying water into same class. So accuracy assessment is done for these three bands and band 7 is giving best result in identification of water. This result is independent of time, method and size of the band image, i.e., for two different years 2013 and 2018 band 7 is having best accuracy. In 2013 the size of band image is larger while the size in 2018 is smaller, even though band 7 is more accurate. In both supervised classification scheme and unsupervised classification scheme band 7 is accurately classifying the data and identifying the water class of both the rivers. Even this result is independent of number of classes used to classify the data; In ISODATA default parameters from ENVI 5.1 (5 to 10 classes) are used while in Minimum Distance we have used only 5 classes. Accuracy of band 7 is better than any other band for identification of water region in multiple sources.

**6. REFERENCES**

1. https://en.wikipedia.org/wiki/Triveni_Sangam

2. T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote sensing and image interpretation*, John Wiley & Sons, fourth edition, 2014, pp 1-52 and 470-568 .

3. J. R. Anderson, E. E. Hardy, J. T. Roach and R. E. Witmer, "A Land Use and Land Cover Classification System for Use with Remote Sensor Data", *United States Government Printing Office*, 1976, pp 1-27.

4. Tri Dev Acharya and Intae Yang," Exploring Landsat 8", International Journal of IT, Engineering and Applied Sciences Research (IJIEASR), Volume 4, No. 4, April 2015, pp 1-21.

5. J. A. Barsi, K. Lee, G. Kvaran, B. L. Markham and J. A. Pedelty, "The Spectral Response of the Landsat-8 Operational Land Imager", *Remote Sens.*, vol.6, No. 10, **2014,** pp 10232-10251.

6. D. Lu and Q. Weng. "A survey of Image Classification Methods and Techniques for Improving Classification Performance." *International journal of Remote sensing* Vol. 28, No. 5, 10 March 2007, pp 823–870.

7. P. M. Mather, and M. Koch, *Computer processing of remotely-sensed images: an introduction*, John Wiley & Sons, Fourth Edition, 2011,  pp 1-26 and 229-283.

8. M. Salah, "A Survey of Modern Classification Techniques in Remote Sensing for Improved Image Classification", *Journal of Geomatics* Vol 11, No. 1 April 2017, pp 1-21.