

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2019; 4(2): 38-43
© 2019 Stats & Maths
www.mathsjournal.com
Received: 10-01-2019
Accepted: 14-02-2019

Jhade Sunil

Department of Agricultural
Statistics and Applied
Mathematics and Computer
Sciences, University of
Agriculture Sciences, GKVK,
Bengaluru, Karnataka, India

SS Patil

Department of Agricultural
Statistics and Applied
Mathematics and Computer
Sciences, University of
Agriculture Sciences, GKVK,
Bengaluru, Karnataka, India

Efficient scheme of classifier on land use pattern

Jhade Sunil and SS Patil

Abstract

Remote sensing is the collection and interpretation of information about an object, area, without being in physical contact with the object. Major Application of remote sensing in field of agriculture are in management of land and water resources, area estimation and monitoring, crop nutrient deficiency detection, soil mapping etc., The scope of the present study is land use and land pattern classification using digital image classification methods, their comparisons and accuracy assessment. The Machine learning algorithms are played vital role to achieve efficient pattern classifications. The supervised classifier is identifying the classes using trained set while in an unsupervised classification the classifier itself develops the spectral classes. Test imagery were obtained through Sentinel-2B Satellite on 15th January 2018. Maximum Likelihood Supervised Classification and Unsupervised Classification were performed using ERDAS 2015 imagine processing. Classification accuracy validations of supervised classes were expressed using confusion matrix. The measures such as F-measure value, Kappa coefficients estimated. The test of significance of the Kappa coefficient was performed using Z- test. Maximum likelihood classification out performed with highest overall accuracy of 72.99 percent. This study helps the farmers and policymakers using early and accurate estimates of yields, managing resources, and estimate area of crop production.

Keywords: Remote sensing, land use and land cover, supervised classification, unsupervised classification, maximum likelihood classification, classification accuracy

1. Introduction

Remote sensing can be defined as the collection and interpretation of information about an object, area, or event without being in physical contact with the object. Every object has its own in nature of unique property of absorbing, reflecting and emitting electromagnetic energy in different bands of spectrum depending on their material type and conditions. Remote sensing imagery has several applications in mapping land-utilize and land cover, agriculture, soils mapping, ranger service, city planning, archeological examinations, military perception, and geomorphological studying, among different uses. For instance, foresters use aerial photos for preparing forest cover maps, finding conceivable access streets, and estimating amounts of trees reaped.

The geographic information system (GIS) is a computer-based tool that allows you to create, manipulate, analyze, store and display information based on its location. GIS makes it possible to integrate different kinds of geographic information, such as digital maps, satellite imageries and global positioning system data (GPS), along with associated tabular database information. GIS and Remote Sensing play significant roles in securing the future of our food production and our population. The importance of food security is directly linked to increases in population density, limitations on agriculture crop yields, and the spread of food deserts. Using GIS, can incorporate all of this information into a single system and execute common database operations.

2. Review of literature

Balak Ram and Kolarkar (1993) [1] have studied Land-use changes in several parts of arid Rajasthan were recognized and mapped on the investigation, semi-detailed and detailed levels using multirate remotely-sensed data, maintained with field check and secondary information's. During the past three years, the net sown area in arid Rajasthan has increased by 36% while current and long fallows have declined by 29% and 41%, respectively.

Correspondence

Jhade Sunil

Department of Agricultural
Statistics and Applied
Mathematics and Computer
Sciences, University of
Agriculture Sciences, GKVK,
Bengaluru, Karnataka, India

The net irrigated area has increased by 140%. Forest and pastures develop highly dishonored though their areas have increased to some amount. Shamsudheen *et al.* (2005) ^[6] have studied land use /land cover mapping for Kumata taluk of Uttar Kannada of Karnataka. The IRS ID LISS III image was used. To perform supervised maximum likelihood classification. The accuracy of classification was evaluated using stratified sampling method. An overall accuracy of 75 percent was obtained. Richards *et al.* (2006) ^[5] have studied Remote sensing digital image analysis provides the non-specialist with an introduction to quantitative estimation of satellite and aircraft derivative remotely retrieved data. It focuses is on those procedures that seem now to have become part of the set of tools frequently used to achieve thematic mapping. Perumal and Bhaskaran (2010) ^[4] have studied that Remote sensing image classification is one among the most significant application in remote sensing. an effective classifier is needed to classify the remote sensing imageries to extract information. They are experimenting with both supervised and unsupervised classification. Here we compare the different classification methods and their performances. It is found that Mahalanobis classifier performed the best in our classification. Patil S.S *et al.* (2014) ^[3] have studied Classification of the Remote sensing satellite imageries color pixels variability of patterns. Machine learning techniques take carried the improved in an accuracy of classification of patterns of features. Challenges in the estimation of various features viz, crop fields, fallow land, buildings, roads, rivers, water bodies, forest, and other trivial items. Urge in the estimation of crop Sunitha abburu and Suresh babu (2015) have studied Satellite image classification process contains grouping the image pixel values into meaningful classes. Satellite image classification techniques can be broadly classified into three categories automatic, manual and hybrid. Majority of the satellite image classification techniques comes under an automatic category. Satellite image classification requirements selection of appropriate classification technique

based on the requirements. Here we compare various researcher's comparative results on satellite image classification methods. Kiran (2017) ^[2] have studied about the classification of satellite imagery. The satellite image was collected through IRS P-6 LISS-IV on 14th November 2015 for Shettikere hobli, Chikkanayakanahalli Taluk. Maximum likelihood classification, Minimum distance to means classification, Mahalanobis distance classification, and Spectral correlation mapper classification and unsupervised classification were performed. Maximum likelihood classification was found to be best with a highest overall accuracy of 74.16 percent followed by minimum distance to mean 70.00 percent, and Unsupervised classification 46.66 percent were observed.

3. Material and methods

3.1. Description of the Study Area

The study area consists of Ambaji Durga Hobli of Chintamani Taluk of chikkaballapur district of Karnataka state. The area lies between 13°24'37.96"N latitude and 78°3'21.64"E longitude.

3.2 Details of image data

Data was taken from sentinel -2B Satellite image of 15th January 2018 is used for the study. The geometrically adjusted images are found from Karnataka state remote sensing application center (KSRSAC) Government of Karnataka, Bengaluru-560097. Sentinel-2B is a European optical imaging satellite that started on 7 March 2017. The satellite holds wide swath high-resolution multispectral imager with 13 spectral bands. It will provide information for farming| and forestry, amongst others allowing for prediction of plants yields. The spatial resolution of the imageries is 10 meters. The images were recorded in three spectral bands, Blue (0.490-0.52μm), Green (0.560-0.58μm), and Red (0.665-0.688μm) and near Infrared (0.842-0.86μm). ArcGIS and ERDAS software used for structures extraction and study.

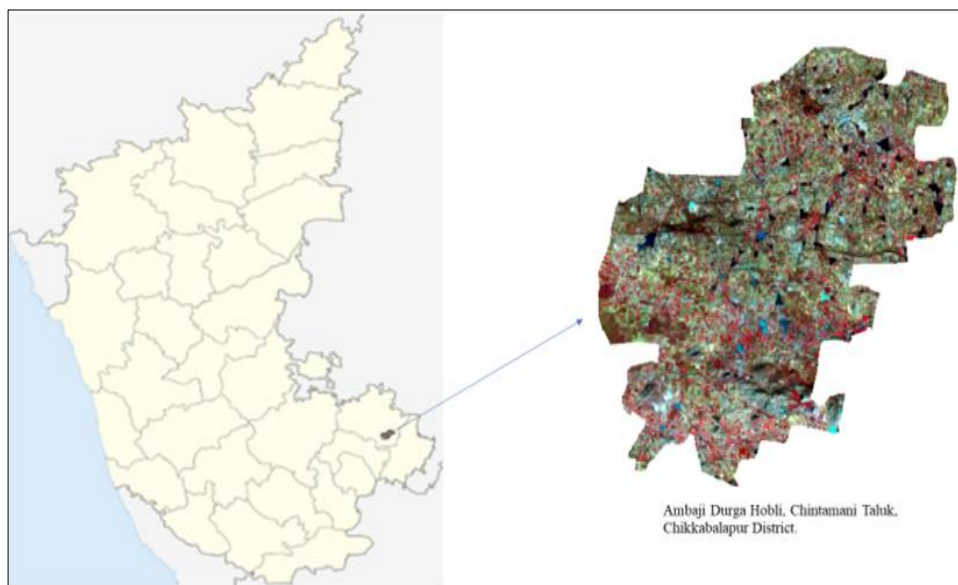


Fig 1: Location map of study area

3.3 Details of Land Use Pattern Classes Considered

In the current study, a broad land use pattern classification system is adopted with eight categories for each training area as follows.

- 1) Agricultural crops
- 2) Horticultural crops

- 3) Grazing land
- 4) Forest
- 5) Water bodies
- 6) Roads
- 7) Build-ups
- 8) Others

3.4 Methods of Image Classification

Image classification is the process of separating the image into diverse areas with some similarities and labelling the regions using additional ground truth information. In the present study, both supervised and unsupervised methods are used for image classification. The supervised classifier is identifying the classes using trained set while in an unsupervised classification the classifier itself develops the spectral classes. In the research two algorithms are being used for a supervised classification and unsupervised classification algorithms.

The Maximum-likelihood algorithm the classification is performed by Variance and Covariance matrix of training set. A digital image in spectral bands are represented by p-matrices of order $M \times N$, where M represents the number of rows and N signifies the number of the number of columns in the image. Every pixel location (i, j) there is a p-dimensional measurement vector.

$X_{ij} = (x_{ij1}, x_{ij2}, x_{ij3}, \dots, x_{ijp})^T$, Representing the digital values in the p- spectral bands. It is assumed that their m land cover classes represented by $\omega_1, \omega_2, \dots, \omega_m$ in the underlying image and the goal of classification is to designate each pixel based on the observations to its class.

3.4.1 Maximum Likelihood Classification Algorithm

Maximum Likelihood Classification is performed, an optional output confidence raster can also be produced. This raster shows the levels of classification confidence. Let $\mu_1, \mu_2, \dots, \mu_m$ and $\Sigma_1, \Sigma_2, \dots, \Sigma_m$ represents the population mean vectors and population variance-covariance matrices for m classes respectively.

The observation vector X_r at pixel r belongs to class c is distributed as a multivariate normal distribution with mean vector μ_c and covariance matrix Σ_c .

Then

$$p_{rc} = \left[\frac{1}{2\pi} \right]^{p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x_r - \mu_c)^t \Sigma_c^{-1} (x_r - \mu_c) \right\}$$

Given the likelihood of pixel r fitting to class c ,

Taking natural log, we have

$$\ln p_{rc} = \frac{p}{2} \ln \left[\frac{1}{2\pi} \right] - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} (x_r - \mu_c)^t \Sigma_c^{-1} (x_r - \mu_c)$$

3.4.2 Unsupervised Classification

Every pixel in an image coordinated to an isolated cluster to find out which set as being most similar to it and it is closet too. Colors are then determined to each class, and the clusters are interpreted after classification based on information of the image or by ground truthing.

3.4.3 Classification of Accuracy Assessment

Classification accuracy is estimated using testing data, i.e., the spatial data consisting of pixels for which the correct classification is known but not used in generating training statistics. Comparison between the classification obtained by the method under consideration and the accurate classification using test data is made, a count of some pixels correctly classified and misclassified are recorded for each class in an error matrix.

The error matrix is a rectangular array of numbers in rows and columns which express the number of pixels assigned to a particular category comparative to the actual category as verified by test data set.

While Kappa coefficient (K) is the measure of agreement of accuracy. It provides a difference measurement between the observed agreement of two maps and agreement that is contributed by chance alone.

The overall accuracy is generally expressed as a percent, with 100% accuracy being a perfect classification where all reference site was classified correctly. Overall accuracy is the easiest to analyze and understand but ultimately only provides the map user and producer with basic accuracy information.

3.4.4 Chi-square (χ^2) test for goodness of fit:

When the data consist of frequencies in discrete categories, the χ^2 test may be used to determine the significance of differences between two independent groups. The null hypothesis is that the two samples of frequencies come from the same population.

The values of χ^2 are distributed approximately as chi-square with $(r-1)(k-1)$ degrees of freedom, where r = number of rows and k = number of columns.

4. Results and discussion

The results are obtained from the satellite image with efficient land use classification by different algorithms and measure the accuracy assessment of classification are presented as following subsections.

1. Collection and digitization of satellite image
2. Classification of pixel data of satellite image with different algorithms
3. Validate the Classification by Kappa Coefficient.

4.1 Collection and digitization of satellite image

To achieve the objectives stated in chapter 1, the satellite image of study area collected from KARSAC which is obtained by remote sensing through the sentinel-2B satellite image on 15th January, 2018. The results obtained by applying the methods as described in the III chapter are conferred below. It is a secondary source data. By using sentinel-2B remote sensing satellite they are collecting the data. Sentinel-2B is the state-of-the-art satellite for the multi-spectral land observation mission with 13 spectral bands in that only 3 are visible infrared bands (red, green, and infrared).

4.2 Classification of data with different algorithms

The basic steps for supervised classification as revealed in chapter III under Section 3.2.1 is as followed. Once the groups of interest are defined sample of homogeneous pixels are selected as training sites of each group by drawing polygons on the false color composite images. These training sites are used to produce statistical descriptors for each land use land cover class. The statistics obtained for the training sites of each class for the study area is presented in the Tables

4.2.1 Maximum Likelihood Classification

Each pixel is classified into training site which one of the land use classes defined in chapter III the resulted classified image is in plate-1. Set of color coding are given to distinguish the corresponding land categories. Table 1 reviled the confusion matrix of Maximum likelihood classification distribution of classes in different categories.

4.2.2 Unsupervised Classification

In this study, unsupervised classification clustering algorithm were used. The Unsupervised classification is performed using 28 clusters and 95 percent confidence threshold, for the study area. The clusters are built, each spectral class is analyzed individually to identify the class type. The spectral

classes belonging to the same class type are combined. Final classification map is obtained for eight land use classes. The classified image is presented in Plate-2 for Ambaji Durga Hobli, Chintamani Taluk. Table 2 revealed the confusion matrix of unsupervised classification distribution of classes in different categories.

Table 1: Classification satellite image result obtained from Maximum Likelihood Classification algorithm for Ambaji Durga Hobli.

Classification categories	Reference categories								Total
	1	2	3	4	5	6	7	8	
1. Agricultural crops	12	.	1	.	.	.	.	.	13
2. Horticultural crops	.	6	.	1	.	.	1	.	8
3. Grazing	1	.	8	.	.	.	.	.	9
4. Forest	2	.	.	13	1	.	.	.	16
5. Waterbodies	.	.	.	.	3	.	.	.	3
6. Built-ups	2	1	.	3	.	17	3	2	28
7. Roads	.	5	1	2	1	7	33	.	49
8. Others	1	.	.	.	.	.	.	9	10
Total	18	12	10	19	6	24	37	11	137



Plate 1: Supervised classification map of Sentinel-2B image of Ambaji Durga Hobli, Chintamani Taluk, chikkaballapur District, using Maximum likelihood Algorithm.

Table 2: Classification satellite image result obtained from unsupervised classification for Ambaji Durga Hobli.

Classification Categories	Reference categories								Total
	1	2	3	4	5	6	7	8	
1. Agricultural crops	8	.	1	.	.	.	4	.	13
2. Horticultural crops	2	9	.	.	.	8	3	.	22
3. Grazing	.	1	4	3	.	.	3	.	11
4. Forest	4	.	1	12	.	.	5	.	22
5. Waterbodies	.	.	.	.	4	.	1	.	5
6. Built-ups	.	1	2	3	1	14	.	.	21
7. Roads	1	1	2	1	1	1	15	1	24
8. Others	3	.	.	.	.	.	6	10	19
Total	18	12	10	19	6	24	37	11	137

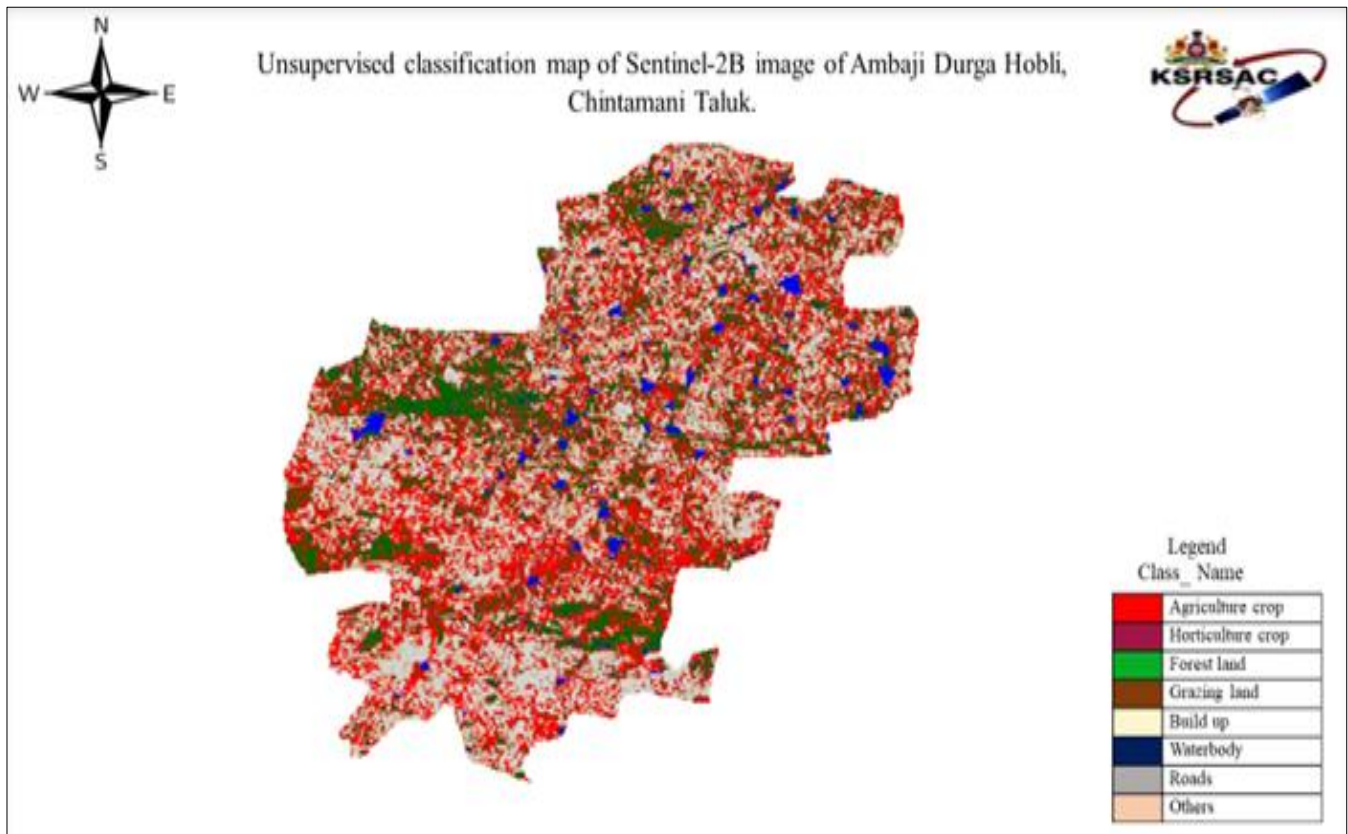


Plate 2: Unsupervised classification map of Sentinel-2B image of Ambaji Durga Hobli, Chintamani Taluk, chikkaballapur District, Karnataka state.

Table 3: F Measures estimated for different classification category using Maximum likelihood classification and unsupervised classification.

Classification Algorithm	Maximum Likelihood Classification			Unsupervised Classification		
	F-measure $F=2rp/r+p$	Producers accuracy (per cent)	User's accuracy (Per cent)	F-measure $F=2rp/r+p$	Producers accuracy (per cent)	User's accuracy (Per cent)
1. Agricultural crops	0.77	66.67	92.31	0.52	44.44	61.54
2. Horticultural crops	0.60	50.00	75.00	0.53	75.00	40.91
3. Grazing	0.84	80.00	88.89	0.39	40.00	36.36
4. Forest	0.74	68.42	81.25	0.59	63.15	54.54
5. Water bodies	0.66	50.00	99.99	0.73	66.66	80.00
6. Built ups	0.65	70.83	60.71	0.63	58.35	66.66
7. Roads	0.73	84.62	64.71	0.50	40.54	62.50
8. Others	0.85	81.82	90.00	0.67	90.90	52.63

4.3 Validate the Classification by Kappa Coefficient

Test of significance is performed for Kappa coefficients of each method. Test of significant difference between Kappa coefficients of different methods. Table 4 shows the test of significance of kappa coefficient at 1 percent level. All these classification algorithms show a variance of kappa value is less than 0.01 it means significance of kappa confident of all

classified algorithms are significance at one percent level for the image.

The validity of classification accuracy was assessed using Kappa statistics which measures the degree of concordance. The Kappa coefficient for all classification significantly differ from random assignment shows Table 4.

Table 4: Test of significance of Kappa coefficient for study area

Classification Algorithm	Kappa (K)	Variance of K	p-Value	Overall accuracy
Maximum Likelihood	0.68	0.00172	< 0.01	72.99
Unsupervised	0.46	0.00213	< 0.01	52.55

Kappa coefficient of maximum likelihood is (0.68) and unsupervised classification algorithms (0.46) are moderate agreement. Overall accuracy Highest overall accuracy in Maximum likelihood classification 72.99 percent lowest overall accuracy found in Unsupervised classification 52.55 percent.

4.3.1 Area estimated with classifications and Ground truth values

The Maximum likelihood classification is achieved the estimation of categories which is found to be more significant on truth observation in case of most of the classes compare to unsupervised classification. Shows Table 5

Table 5 Area estimated with classifications and Ground truth values.

Classes	MLE	Unsupervised classification	Ground truth
Agricultural crops	3376.28	3621.76	3418.73
Horticultural crops	2810.30	2492.83	2994.16
Grazing	895.91	1439.08	938.69
Forest	2063.27	2318.26	2142.52
Water bodies	1270.33	623.75	1300.09
Built-ups	593.65	781.32	614.56
Roads	968.83	913.64	721.62
others	2925.48	2713.36	2774.25

4.3.2 Chi-square (χ^2) test for goodness of fit

Table 6: Test of significance of chi-square for Ambaji Durga Hobli satellite image

Classifiers	Chi square value	p- value
Maximum Likelihood	520.12**	2.02E-16
Unsupervised	207.98**	1.32E-54

** : significant at 1 percent level.

In this technique we are comparing the significant difference between the ground truths frequencies with the frequencies obtained with the two classification algorithms to each result is presented in below table.

Table 6 shows a test of significance of chi-square for the study area of the satellite image maximum likelihood classification algorithm nearer to the ground truth value. It found that highest probability (2.02E-16) and unsupervised classification shows (1.32E-54) of p-values.

5. Conclusion

Maximum likelihood classification algorithm on satellite imageries out performs in statistically significantly converging to actual classes. Kappa coefficient derives the significantly different from a random assignment. The classifier achieved more than 95% classification accuracy in agricultural crops. Unsupervised classification methods govern spectral classes automatically, but generally, show the restricted capacity to precisely partition the landscape into normal classes. Maximum likelihood classification algorithm is observed to be best with a highest overall accuracy of 72.99 percent for a study area. It classified is based on variance-covariance matrices for each class. Unsupervised classification depends on the purpose of study and condition of the selected area. It has an overall accuracy of 52.55 percent for the study area. For validation purpose, we use chi-square test for goodness of fit. Both the tests are found that Maximum likelihood is very near to the ground truth observation near to the ground truth observation. Higher accuracy would help to increasing yields, managing resources, reduction of input cost and prediction of future outcomes and also policymaker to estimate crops yield and area before harvest.

6. The future line of work

The study can be extended to large area say for taluk, districts, classification according to the land use and land cover status, was helped to control and supervision the main quantity about land use types and the implementation of the plans and also it helps to the policymaker to estimate crops yield and area before harvest. Agriculture industry also using GIS techniques for increasing yields, managing resources, reduction of input cost and prediction of future outcomes.

7. References

1. Balak Ram, Kolarkar AS. Remote sensing application in monitoring land-use changes in arid Rajasthan. Int. J Remote sens. 1993; 14(17):3191-3220.
2. Kiran KS. Pattern classification scheme for remote sensing satellite imagery. M. Sc. (Agri.) Thesis (Unpub.), Univ. Agric. Sci., Bengaluru, 2017.
3. Patil SS, Sachidanad UB, Angadi, Prabhuraj DK. Machine Learning Technique Approaches verses Statistical methods in Classification of multispectral remote sensing data using Maximum Likelihood Classification. Int. J Adv. Remote Sensing and GIS. 2014; 3(1):525-531.
4. Perumal K, Bhaskaran R. Supervised Classification Performance of Multispectral Images. J of computing. 2010; 2:2151-9617.
5. Richards JA. Remote Sensing Digital Image Analysis, Springer-Verlag Berlin, 2006, 240.
6. Shamsudheen M, Dasog GS, Tejaswini B. Land use /land cover mapping in the coastal area of North Karnataka using remote sensing. J Indian Soci. Of Remote Sens. 2005; 33(1):155-1.