

Original Research Article

<https://doi.org/10.20546/ijcmas.2019.810.304>

Evaluation of the Performance of Supervised Classification Algorithms in Image Classification

Jhade Sunil* and Abhishek Singh

Department of Farm Engineering (Agricultural Statistics), Banaras Hindu University,
Varanasi, India

*Corresponding author

ABSTRACT

Keywords

Remote sensing;
Land use pattern;
Supervised
Classification;
Classification
Accuracy; kappa
coefficient

Article Info

Accepted:
18 September 2019
Available Online:
10 October 2019

This study presents a land use pattern classification of satellite imagery. The Machine learning algorithms are overseen to pattern classifications. The supervised classifier is identifying the classes using trained set. Compiled classification has to be improvised using efficient algorithms with appropriate threshold values. The statistical significance of satellite image classifies into essential classes is of greater importance in remote sensing pattern classification methods. Test imagery were obtained through Sentinel-2B Satellite on 15th January 2018 for Ambaji Durga Hobli, Chikkaballapur District. Maximum Likelihood Classification, Minimum Distance to means Classification, Mahalanobis Distance Classification, Spectral Correlation Mapper Classification were performed using ArcGIS 10.5.1 and ERDAS 2015 imagine image processing soft wares. Accuracy of the classification expressed using confusion matrix. The measures such as overall accuracy, F-measure value, Kappa coefficients its variance were 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 per cent followed by Minimum distance to mean 68.61 per cent, Mahalanobis distance 61.31 per cent, Spectral correlation mapper 56.20 per cent This study helps the farmers using early and accurate estimates of yields, estimate area of crop production.

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. Remote sensing of nature by geographers is generally finished with the assistance of mechanical devices known as sensors. These contraptions have an incredibly enhanced capacity to get and record data around a protest with no

physical contact. Regularly, these sensors are situated far from the question of enthusiasm by utilizing helicopters, planes, and satellites. Sensors depends on the property of the material (auxiliary, substance, and physical), surface coarseness, an angle of incidence, intensity, and wavelength of radiant energy.

The geographic information system (GIS) is a system of hardware, software, and procedures

to facilitate the management, manipulation, analysis, modeling, representation, and display of geo referenced data to solve complex issues regarding planning and management of resources. In supervised classification, spectral features a few regions of known land use types are removed from the image. These areas are known as the training areas. Each pixel in the entire image is then classified as belonging to one of the classes depend upon how shut its spectral highlights are to the spectral features of the training areas. Surely understood example: minimum distancemean classification, Maximum Likelihood classification, Mahalanobis distance classification, and spectral correlation mapper. Application of remote sensing and GIS in agriculture was identification, area estimation and monitoring, crop nutrient deficiency detection, soil mapping, crop condition assessment, reflectance modeling, crop yield modeling and production forecasting.

Shamsudheen *et al.* (2005) 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. The overall accuracy of 75 percent was obtained. Sharma and Leon (2005) was studied on the supervised classification using Maximum likelihood algorithm on three dates of IRS LISS3 satellite data identify the outcome of seasonal spectral variation on land use land cover.classification for the study area falling in the Sloan district of Himachal Pradesh state.It was found that summer data set was better with overall accuracy 76 percent as relating to winter and spring dataset with classification accuracy 49 percent and 46 percent respectively. Madhura and Venkatachalam (2013) have a classification of different land use land cover categories from the raw satellite image using supervised

classifiers and performances of the classifiers are studied. Classification is performed based on the spectral features using Maximum likelihood classification algorithm, Minimum distance to mean classification algorithm, and Mahalanobis classification. Maximum likelihood produced the 93.33% overall efficiency and minimum distance showed the overall classification accuracy of 85.72% and Mahalanobis gave the overall accuracy of 90.00%. Patil *et al.* (2014) have studied Classification of the Remote sensing satellite imageriesarecolor pixels variability of patterns. Machine learning techniques take carried the improved in 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. The study achieved more than 95% classification accuracy in agricultural crops. Manish and Rawat (2015) have studied the Digital change identification techniques by using multi-temporal satellite imagery helps in understanding landscape dynamics. The Supervised classification methodology has been employed using maximum likelihood technique in ERDAS 9.3 Software. Image categorized into five different classes namely vegetation, agriculture, barren, built-up and water body. Accuracy assessment of the land use classification results obtained showed an overall accuracy of 90.29 percent for 1990 and 92.13 percent for 2010. The Kappa coefficients for 1990 and 2010 maps were 0.823 and 0.912 respectively.

Materials and Methods

Description of the Study Area

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

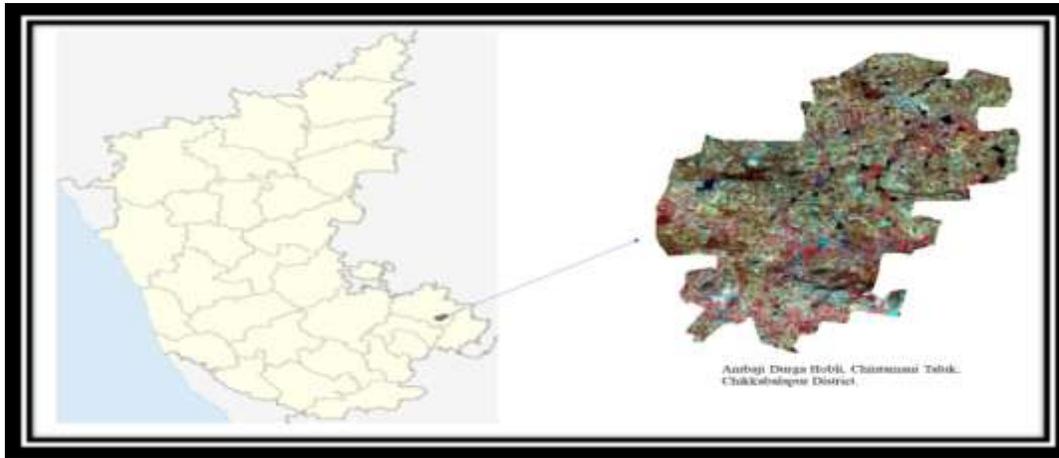


Figure.1 Location map of study area

Details of image data

Data was taken from sentinel -2B Satellite image of 15th January 2018 is used for the study. The image collected from Karnataka state remote sensing application center (KSRSAC) Government of Karnataka, Bengaluru-560097. Sentinel-2B is a European optical imaging satellite. The satellite holds wide swath high-resolution multispectral imager with 13 spectral bands. The spatial resolution of the imageries is 10 meters. The images were recorded in three spectral bands, Blue (0.490-0.52 μ m), Green (0.560-0.58 μ m), and Red (0.665-0.688 μ m) and near Infrared (0.842-0.86 μ m). ArcGIS and ERDAS software used for structures extraction and study.

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

9) Others

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, four supervised classification methods namely Maximum-likelihood algorithm, Mahalanobis distance algorithm, Minimum distance to means algorithm, spectral correlation Mapper, are used for image classification. Supervised classification is the technique most often used for the quantitative analysis of remote sensing image data. At it is the core concept of segmenting the spectral domain into regions that can be associated with the ground cover classes of interest to a particular application.

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 \mathbf{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

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

Given the likelihood of pixel r fitting to class c ,

Taking natural log, we have

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

Minimum Distance to means Classification Algorithm

The minimum distance to mean classifier is simplest mathematically and very efficient in computation. When the number of training samples per class is limited, it can be more real to option to a classifier that does not make use of covariance information but then instead depends only upon the mean positions of the spectral classes, noting that for a given number of samples these can be more accurately estimated than covariances. The so-called minimum distance classifier, the most used distance calculation method is Euclidean distance.

Symbolically let $\omega_1, \omega_2, \dots, \omega_m$ indicate the m land cover classes in the image with unknown mean vectors, $\mu_1, \mu_2, \dots, \mu_m$. Let $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_m$ represent the sample mean vectors of the m classes estimated from the training set.

$$\bar{X}_c = (\bar{x}_{c1}, \bar{x}_{c2}, \dots, \bar{x}_{cp})^T$$

where mean is calculated over all pixels in the training set of class c , for $c = 1, 2, \dots, m$ classes and $k=1, 2, \dots, S$ denote the digital value of r th pixel. $X_r = (X_{r1}, X_{r2}, \dots, X_{rp})^T$.

Let D_{rc} denote the Euclidean distance between

pixel r and the class c ,

Then,

$$D_{rc} = \sqrt{\sum_{k=1}^p (x_{rk} - \bar{x}_{rk})^2}$$

For all $c=1, 2, \dots, m$

The minimum distance to means classifier assigns pixel r to class c if

$$D_{rc} \leq D_{rq}$$

For all $q= 1, 2, \dots, m$ classes, $q \neq c$

Classification using Mahalanobis Distance Algorithm

The Mahalanobis distance originally refers to a distance measure that incorporates the correlation among the features.

Let $\mu_1, \mu_2, \dots, \mu_m$ and $\Sigma_1, \Sigma_2, \dots, \Sigma_m$ denotes the population mean vectors and population variance-covariance matrices for m classes respectively.

The observations vector \mathbf{X}_r at pixel r when it belongs to class c is distributed as a multivariate normal distributed with mean vector μ_c and covariance matrix Σ_c .

Then,

$$D_{rc}^2 = (\mathbf{X}_r - \mu_c)^T \Sigma_c^{-1} (\mathbf{X}_r - \mu_c)$$

Spectral Correlation Mapper Classification Algorithm

The Spectral Correlation Mapper (SCM) method is imitative of Pearson Correlation Coefficient that removes negative correlation and maintains the Spectral Angular Mapper (SAM) characteristic of minimizing the shading effect resultant in better results. The SCM varies from -1 to 1 and SAM varies from 0 to 1 .

The SCM algorithm method, similar to SAM, uses the reference spectrum as defined by the

investigator. SCM presents the following formula.

$$R = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

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.

$$\hat{K} = \frac{P_o - P_e}{1 - P_e}$$

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.

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.

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.

Collection and Classification of satellite image with different algorithms

Validate the Classification by Kappa Coefficient.

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.

Maximum Likelihood Classification

Each pixel is classified into training site which one of the land use classes defined in chapter III. Table 4.1 revealed the confusion matrix of Maximum likelihood classification distribution of classes in different

categories. Table 4.5 revised F Measures estimated for different classification.

Minimum Distance to Means Classification

Each pixel classified into one of land use land cover classes described in chapter III Table 4.2 shows the confusion matrix of Minimum distance to mean classification distribution of classes in different classes. Table 4.5 revised F Measures estimated for different classification.

Mahalanobis Distance Classification

Every single pixel is classified into any one of

the land cover classes described in chapter III Table 4.3 shows the confusion matrix of Mahalanobis distance classification distribution of classes in different categories. Table 4.6 revised F Measures estimated for different classification.

Spectral Correlation Mapper

The confusion matrix of Spectral Correlation mapper algorithm shows in table 4.4. Each pixel in that image is classified as any one of the land use class. Spectral correlation mapper described in chapter III. Table 4.6 revised F Measures estimated for different classification.

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. Water bodies	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

Table.2 Classification satellite image result obtained from Minimum Distance to Means Classification algorithm

Classification Categories	Reference categories								Total
	1	2	3	4	5	6	7	8	
1. Agricultural crops	11	1	.	1	13
2. Horticultural crops	3	10	.	3	.	1	2	.	19
3. Grazing	.	.	7	1	.	.	1	1	10
4. Forest	.	.	.	9	9
5. Waterbodies	.	.	.	1	5	.	.	.	6
6. Built-ups	1	.	.	2	1	18	5	.	27
7. Roads	1	.	1	.	.	.	26	2	30
8. Others	2	1	2	2	.	5	3	8	21
Total	18	12	10	19	6	24	37	11	137

Table.3 Classification satellite image result obtained from Mahalanobis Distance Classification algorithm

Classification Categories	Reference categories								Total
	1	2	3	4	5	6	7	8	
1. Agricultural crops	10	3	.	13
2. Horticultural crops	.	4	1	1	6
3. Grazing	1	.	5	6
4. Forest	.	2	.	14	16
5. Waterbodies	3	.	.	.	3
6. Built-ups	2	1	3	3	1	18	5	.	33
7. Roads	5	5	2	2	2	6.	27	7	56
8. Others	1	3	4
Total	18	12	10	19	6	24	37	11	137

Table.4 Classification satellite image result obtained from Spectral Correlation Mapper

Classification Categories	Reference categories								Total
	1	2	3	4	5	6	7	8	
1. Agricultural crops	13	.	1	2	1	.	2	.	19
2. Horticultural crops	.	6	1	1	.	.	1	2	11
3. Grazing	.	.	4	.	.	1	.	.	5
4. Forest	2	2	1	12	.	.	4	.	21
5. Waterbodies	3	1	1	1	6
6. Built-ups	1	3	1	2	.	15	5	2	29
7. Roads	1	.	2	2	2	5	19	1	32
8. Others	1	1	.	.	.	2	5	5	14
Total	18	12	10	19	6	24	37	11	137

Table.5 F Measures estimated for different classification category using Maximum likelihood classification and Minimum Distance to Means classification.

Classification Algorithm	Maximum Likelihood Classification			Minimum Distance to Means		
	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	0.71	61.11	84.62
2. Horticultural crops	0.60	50.00	75.00	0.65	83.33	52.63
3. Grazing	0.84	80.00	88.89	0.70	70.00	70.00
4. Forest	0.74	68.42	81.25	0.64	47.37	99.99
5. Water bodies	0.66	50.00	99.99	0.83	83.33	83.33
6. Built ups	0.65	70.83	60.71	0.72	75.00	69.23
7. Roads	0.73	84.62	64.71	0.77	70.28	86.25
8. Others	0.85	81.82	90.00	0.47	72.73	34.78

Table.6 F Measures estimated for different classification category using Mahalanobis Distance classification and Spectral Correlation Mapper classification.

Classification Algorithm	Mahalanobis Distance Classification			Spectral Correlation Mapper		
	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	0.65	55.56	76.92	0.70	72.22	68.42
2. Horticultural crops	0.44	33.34	66.67	0.52	50.00	54.55
3. Grazing	0.62	50.00	83.33	0.54	40.00	80.00
4. Forest	0.80	73.68	87.50	0.60	63.16	57.14
5. Water bodies	0.66	50.00	99.99	0.50	50.00	50.00
6. Built ups	0.63	75.00	54.55	0.57	62.50	51.73
7. Roads	0.59	72.97	48.21	0.55	51.35	59.35
8. Others	0.40	27.28	75.00	0.40	45.45	35.71

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 4.7 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. Overall accuracy of all supervised classification shows Table 4.7.

Table.7 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
Minimum distance to means	0.63	0.00184	< 0.01	68.61
Mahalanobis Distance	0.52	0.00204	< 0.01	61.31
Spectral Correlation Mapper	0.48	0.00210	< 0.01	56.20

Kappa coefficient of maximum likelihood is (0.68), Minimum distance to means (0.63), Mahalanobis Distance (0.52) and Spectral Correlation Mapper (0.48). Overall

accuracy Highest overall accuracy in Maximum likelihood classification 72.99 percent lowest overall accuracy found in Spectral Correlation Mapper 56.20 percent.

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.8.

Table.8 Area estimated with classifications and Ground truth values.

Classes	Maximum likelihood	Minimum distance to means	Mahalanobis distance	Spectral correlation mapper	Ground truth
Agricultural crops	3376.28	3183.63	2864.19	2643.29	3418.73
Horticultural crops	2810.30	2642.70	3165.42	2294.68	2994.16
Grazing	895.91	794.54	568.51	1263.04	938.69
Forest	2063.27	2591.38	2663.72	3568.92	2142.52
Water bodies	1270.33	1125.75	980.48	1129.73	1300.09
Built-ups	593.65	926.63	1139.78	873.48	614.56
Roads	968.83	1237.29	432.79	582.52	721.62
others	2925.48	2402.16	3089.39	2548.34	2774.25

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

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.9 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
Minimum Distance	441.86**	1.92E-05
Mahalanobis Distance	345.13**	1.60E-12
Spectral Correlation Mapper	257.11**	1.38E-68

** : significant at 1 percent level.

Table 4.9 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), next highest probability (1.92E-05) observed in a minimum distance to mean classification, Mahalanobis distance classification shows (1.6E-12), spectral correlation mapper shows (1.38E-68) p-value.

In Agriculture, satellites have a capability to image individual fields, regions and counties on a frequent reenter cycle. Digital image analysis is a vital role in remote sensing area like land use land cover classification. The spectral response of a particular land cover class deviates from its ideal response due to the presence of noise. Application of statistical methods in remote sensing image classification in order to partition the noisy

image into its constituent classes is of great importance. Maximum likelihood classification algorithm is observed to be best with a highest overall accuracy of 72.99 percent for a study area. Maximum likelihood algorithm is a parametric method of Classification which depend on the Gaussian probability model for each class. It classified is based on variance-covariance matrices for each class, In case of Minimum distance to mean classification which makes use of mean vectors of training sets to assign an unknown pixel to the category using Euclidean distance, the overall accuracy of 68.61 percent attained for the study area, it is standing next to Maximum Likelihood. Mahalanobis distance classification which considers mean vectors and population variance-covariance matrices for each class, the overall accuracy of 61.31 percent of the study area, stands next to maximum likelihood and minimum distance to mean classification. Spectral correlation mapper it is based on targeted spectrum and reference spectrum, the overall accuracy 56.20 percent.

The future line of work:

The maximum likelihood classification is superior method among classifiers, for of land areas and production estimates through classification procedures. 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.

References

- Madhura, Suganthi. and Venkatachalam., 2013, Comparison of supervised classification methods on remote sensed satellite data an application in Chennai, India. *International journal of science and research*, 6(14):128-143.
- Manish, Kumar. and Rawat, J.S., 2015, Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarkhanda, India, *The Egyptian J. Remote Sensing and Space Sci.*, 18(5):77-84.
- Patil, S.S., Sachidanad, U.B., Angadi and Prabhuraj, D. K., 2014, 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*.3(1): 525-531.
- Shamsudheen, M., Dasog, G. S. and Tejaswini, B., 2005, Land use /land cover mapping in the coastal area of North Karnataka using remote sensing. *J. Indian Soci. Of Remote Sens.*, 33(1):155-163.
- Sharma, D.P. and Bren, Leon., 2005, Effect of seasonal spectral variation on land cover classification, *J. Indian Soci. Of Remote sens.*, 33(2):203-209.

How to cite this article:

Jhade Sunil and Abhishek Singh. 2019. Evaluation of the Performance of Supervised Classification Algorithms in Image Classification. *Int.J.Curr.Microbiol.App.Sci.* 8(10): 2634-2643. doi: <https://doi.org/10.20546/ijcmas.2019.810.304>