Cloud Publications
International Journal of Advanced Remote Sensing and GIS
2014, Volume 3, Issue 1, pp. 525-531, Article ID Tech-249
ISSN 2320 - 0243

Research Article **Open Access**

Machine Learning Technique Approaches Versus Statistical Methods in Classification of Multispectral Remote Sensing Data using Maximum Likelihood Classification: Koluru Hobli, Bellary Taluk, District, Karnataka, India

S.S. Patil[1], Sachidananda[1], U.B. Angadi[2], and D.K. Prabhuraj[3]

[1]Department of Computer Science, University of Agricultural Sciences, Bangalore, Karnataka, India
[2]CABin, ISRI, New Delhi, India
[3]KSRSAC, Bangalore, Karnataka, India

Correspondence should be addressed to S.S. Patil, spatilsuasb@gmail.com

Publication Date: 28 April 2014

Article Link: http://technical.cloud-journals.com/index.php/IJARSG/article/view/Tech-249

Abstract Classification is challenging task on complex features of Remote sensing satellite imageries color pixels variability of patterns. Machine learning techniques have delivered the improved in accuracy of classification of patterns of features. Remote sensing color based imageries having hard to cluster color pixels with variability in intensity of colors. Challenges in estimation of various features viz, crop fields, fallow land, buildings, roads, rivers, water bodies, forest, and other trivial items. Urge in estimation of crop yield predictions through satellite imageries. We are attempted to converging accuracy of estimation of vegetation crop yield of fields. Kappa coefficient to achieve high degree accuracy estimation of crop wise with suitable thresh hold to ground truth data.
Keywords *Machine Learning Techniques; Supervised Classification; Maximum Likelihood Classification; Kappa Coefficient; Classification Accuracy; f-measure*

1. Introduction

Remote sensing is the science and art of obtaining information about an object, geographic area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation. This study deals with remote sensing data acquired through earth observation satellites. Remote sensing image analysis is done to extract useful information about the earth surfaces. An important step in the analysis of such data is the process of land cover classification. In this the image pixels are assigned to different land cover classes based on the spectral measurements of each pixel. Pattern recognition technique such as Maximum likelihood classification is followed in this process. Machine learning techniques can be applied in a supervised and unsupervised manner. Supervised Classification is a procedure for identifying spectrally similar areas on an image by identifying 'training' sites of known targets and then extrapolating those spectral signatures to other areas of unknown targets. Classification of images based remote sensing have thrust research area on of the remote sensing community on feasibility study on environmental and socioeconomic applications are based on the classification results (Lu and Weng, 2007). The efficient evaluation of amount of changes in land use land cover through remote sensing satellite digital data analysis make decision support to develop effective plans for the management of land (Gordon, 1980). The assessment of technology of remote sensing has instigated it to develop one of the most regularly used methods in the earth. Supervised classification has been used in this study. Supervised classification of multispectral remote sensing imagery is commonly used for land use land cover determination (Duda and Canty, 2002). There is a consistent logic to all of the supervised classification routines in almost all image processing softwares. In addition, there is a basic sequence of operations that must be followed no matter which of the supervised classifiers is used. In this study the following sequence of operations were used. In thematic mapping from remotely sensed data, the term accuracy is used typically to express the degree of 'correctness' of a map or classification (Foody, 2002). The overall accuracy is expressed as a percentage of the test pixels successfully assigned to the correct classes. Maximum Likelihood produced the highest accuracy with overall accuracy of 87%. Then followed by Mahalanobis gave the overall classification accuracy of 77% and Minimum distance showed the overall classification accuracy of 71%.

2. Materials and Methods

2.1. Description of the Study Area

Study area consists of Koluru Hobli of Bellary Taluk and District of Karnataka, which lies between $15^0 9'$ to $15^0 15'$N latitude and $76^0 55'$ to $76^0 92'$E longitudes.

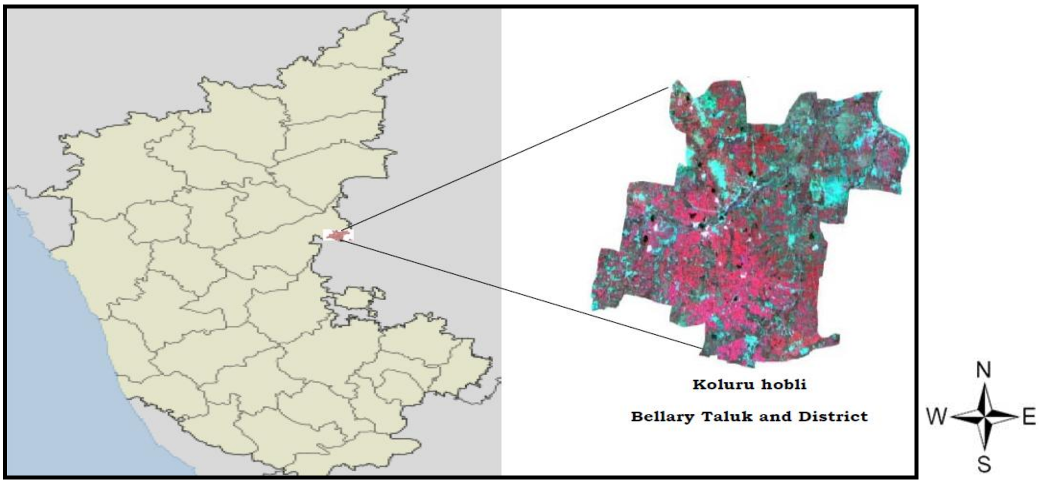


Figure 1: *Sattillite Imagieries of Koluru Hobli of Bellary Taluk and District of Karnataka State*

2.2. Details of Image Data

IRS (Indian Remote Sensing Satellite)-P6 LISS-III (Linear Imaging Self Scanner) imageries of 21st November, 2010 is used for the study. The geometrically corrected imageries are obtained from National Remote Sensing Agency, Department of Space, Government of India, Hyderabad. The topographical map of the study area is over laid on this image to extract the digital image of the study area (Plate–2). The spatial resolution of the images is 23.5 mt. The imageries were recorded in three spectral bands. Among these only the first two namely Green, Red are in the small range of electromagnetic spectrum and third one namely blue band is useful in identification of green vegetation like crop. Ground truth data collected during field visits in the study area and the toposheets are used to accomplish the task of selection of training sites for each category for training the classifier in supervised classification. A part of the data was used as test sites for assessing classification accuracy.

2.3. Details of Land Cover Classes Considered

The categories of interest were carefully selected and defined to successfully perform digital image classification. In the present study a broad land use/land cover classification system is adopted with six categories for each study area as follows.

Land use/Land cover categories of Koluru Hobli, Bellary Taluk.

1) Paddy
2) Cotton
3) Jowar
4) Chilly
5) Fallow land
6) Water bodies

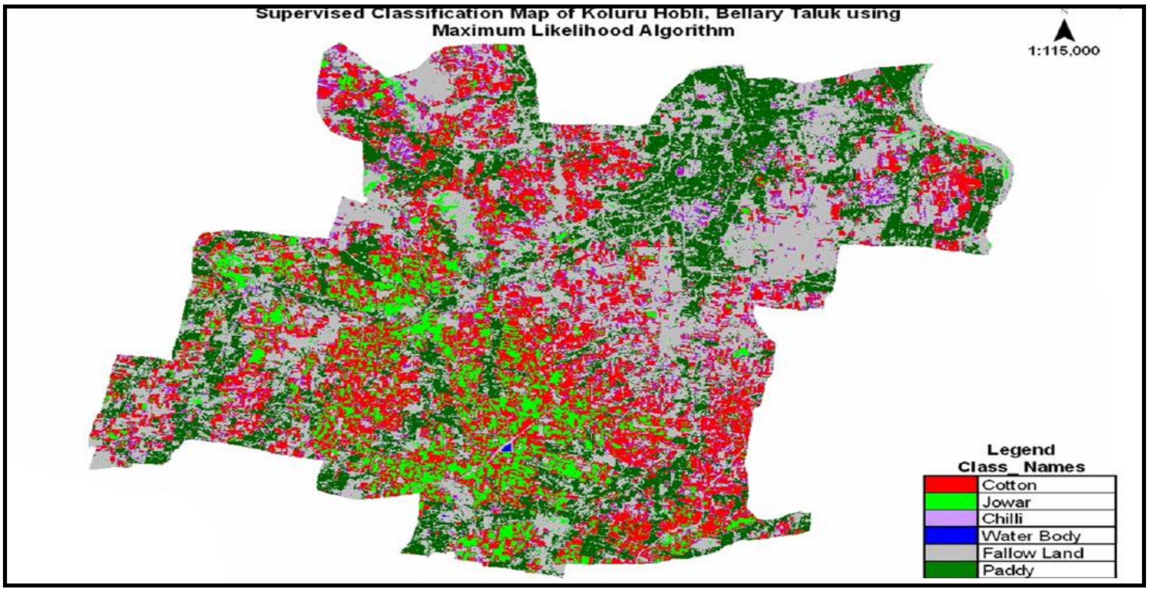


Figure 2: Supervised *Classification Map of Koluru Hobli, Bellary Taluk and District, Karnataka State*

2.4. Methods of Image Classification

Image classification is the process of dividing the image into different regions with some similarly and labeling the regions using supplementary ground truth information. In the present study both supervised and unsupervised methods are used for image classification. All classification are done using ERDAS imagine 9.1 software at the Karnataka State Remote Sensing Application Center, Department of IT and BT, Government of Karnataka, M.S. Building, Dr. Ambedkar Veedhi, Bangalore - 01. Mahalanobis distance classifier is similar to Minimum Distance, except that the covariance matrix is used in the equation. This algorithm assumes that the histograms of the bands have normal distributions (Perumal and Bhaskaran, 2010). Var. and Covariance are figured in so that clusters that are highly varied lead to similarly varied classes and vice versa. The Minimum distance classifier is based on training site data. This classifier characterizes each class by its mean position on each band. Minimum distance classifier is highly recommended in all image classification applications (Richards, 1995). The classification is performed by placing a pixel in the class of the nearest mean. The minimum distance algorithm is also more attractive since it is a faster technique than the maximum likelihood classification.

2.5. Maximum Likelihood Classification Algorithm

Maximum Likelihood Classifier (Gaussian): A statistical decision rule that examines the probability function of a pixel for each of the classes, and assigns the pixel to the class with the highest probability. In this algorithm each pixel is modeled to have multivariate normal distribution. Let $\mu_1, \mu_2, \ldots, \mu_m$ and $\Sigma_1, \Sigma_2, \ldots, \Sigma_m$ denote the population mean vectors and population variance – covariance matrices for m classes respectively. The observation vector $\mathbf{X}_r$ at pixel r when it belongs to class c a multivariate normal distribution with mean μ_c and covariance matrix Σ_C.

Then,

$$P_{rc} = \left[\frac{1}{2\Pi}\right]^{P/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(X_r - \mu_c)' \Sigma_c^{-1}(X_r - \mu_c)\right\}$$

Gives the likelihood of pixel r belonging to class c

Taking logarithm,

$$\ln P_{rc} = \frac{p}{2} \ln\left[\frac{1}{2\Pi}\right] - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(X_r - \mu_c)' \Sigma_c^{-1}(X_r - \mu_c)$$

Ignoring $\frac{p}{2} \ln\left[\frac{1}{2\Pi}\right]$ which is a constant, the maximum likelihood algorithm assigns pixel r to class c if and only if

$\ln P_{rc} \geq P_{rq}$, for all q= 1, 2, ---, m classes, q≠c.

Since the class mean vectors μ_c and covariance matrix Σ_C are unknown, the sample estimates are obtained from the training set.

Let $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_m$ be the sample mean vectors and $V_1, V_2 \ldots V_m$ be the sample variance – covariance matrices estimated from the training data for m classes respectively.

The pixel assignment is made based on the estimated value of p_{rc}

$$\widehat{p_{rc}} = [-0.5 \ln[\det(V_c)] - [0.5(X_r - \bar{X}_c)^T (V_c)^{-1}(X_r - \bar{X}_c)$$

Where, X_r is the observation vector for unclassified pixel r.

The pixel is assigned to that class for which it has the highest similarity if being a member. The decision rule is given as, assign pixel r to the class c if, and only if,

$\widehat{p_{rc}} \geq \widehat{p_{rq}}$ for all q= 1, 2, ---, m classes, q≠c.

The classification of the whole image is performed on a pixel by pixel basis. Every pixel is assigned to one of the mutuality exclusive classes based on the likelihood as described above and no pixel remains unclassified.

2.6. Classification Accuracy

The extent to which a manual or automatic processing system correctly identifies selected classes.

Kappa Coefficient

A statistical measure of the agreement, beyond chance, between two maps (e.g. output map of classification and ground-truth map). It is represented by the symbol kappa hat or k hat.

Error Matrix

An error matrix is a square array of numbers set out in rows and columns which express the number of sample units (pixels) assigned to a particular category relative to the actual category as verified by test data set.

3. Results and Discussion

In the classification phase, supervised classification technique was chosen to classify the images. The technique Maximum Likelihood was performed to the images. Accuracy assessment was approved out to compute the probability of error for the classified map. An overall randomly sample points were chosen for accuracy assessment. The measures of accuracy were tested in this study viz overall accuracy, confusion or error matrix and kappa coefficient. Several measures of classification accuracy may be derived from a confusion matrix. Kappa coefficient was generated to portray the amount of agreement between the classification effect and the validation sites after random agreements by possibility are detached from consideration these data.

Table 1: *Producers and Users Accuracy*

Classified Category	Producer's Accuracy (Percent)	User's Accuracy (Percent)
Jowar	93.33	82.35
Cotton	88.00	91.66
Chilly	95.00	90.47
Fallow Land	90.00	93.10
Paddy	90.00	95.74
Water body	93.33	90.32

Kappa coefficient and overall accuracy results of the measures of accuracy are shown in the Table 1. The overall accuracy is expressed as a percentage of the test pixels successfully assigned to the correct classes. Maximum Likelihood produced the highest accuracy with overall accuracy achieved 91%. The user's accuracy and producer's accuracy for individual classes ranged between 88.00 to 95.00 percent and 82.35 to 95.74 percent for different categories using maximum likelihood algorithm. Producers and users accuracy obtained and presented in the Table 1 for all the classes under study.

3.1. Test of Significance of Kappa Coefficients for Koluru Hobli, Bellary Taluk

The kappa coefficient is found to be 0.8321 and the variance of kappa is found to be 0.000675. The Kappa coefficient for Maximum likelihood classification was highly significant, implying that classifier produced classification significantly different from a random assignment.

3.2. Area Estimates for Koluru, Hobli, Bellary Taluk (Hectares) using Maximum Likelihood Classification

The estimated area for all the classes under study using Maximum likelihood classification is presented in the Table 2, and which is found to be more nearer to the ground truth observation.

Table 2: *The Estimated Area for all the Classes under Study Using Maximum Likelihood Classification*

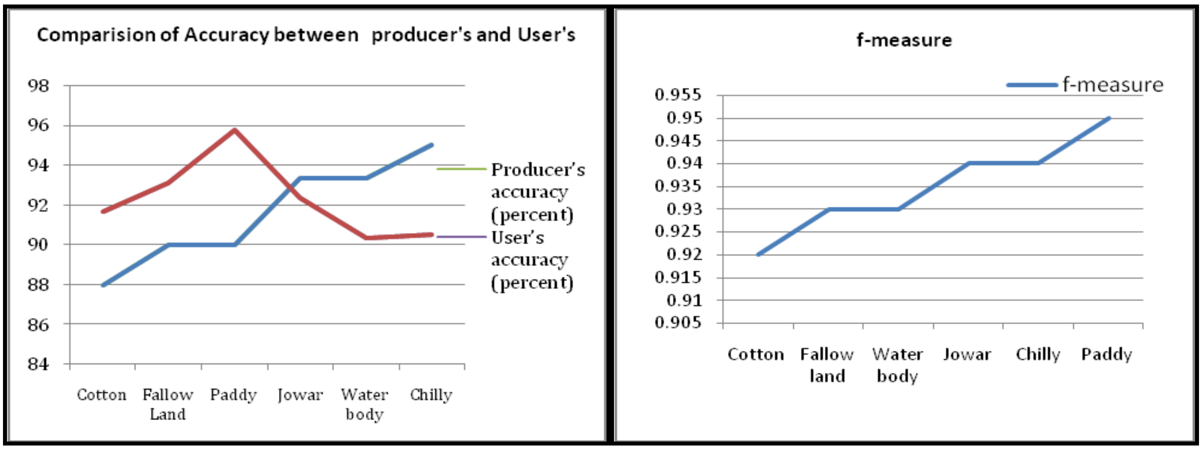
Sl. No.	Classification Categories	Max-Like Classification (ha)
1	Jowar	2209.32
2	Cotton	6590.27
3	Chilly	2800.49
4	Fallow land	14744.87
5	Paddy	7245.33
6	Water body	65.32

The Maximum likelihood classification is achieved the estimation of categories which is found to be more significant on truth observation.

Table 3: *f Measure Table*

Sl. No.	Classification Categories	Recall	Precision	f-measure
1	Jowar	0.93	0.95	0.94
2	Cotton	0.93	0.90	0.92
3	Chilly	0.94	0.94	0.94
4	Fallow land	0.92	0.93	0.93
5	Paddy	0.97	0.93	0.95
6	Water body	0.92	0.93	0.93
	Total f-measure	**0.94**	**0.93**	**0.94**

The measure of accuracy was tested in this study namely overall accuracy, confusion or error matrix and kappa coefficient. Many measures of classification accuracy may be derived from a confusion matrix. Kappa coefficient was generated to describe the proportion of agreement between the classification result and the validation.



(a) **(b)**

Figure 3: *(a) Comparision of Accuracy between producer's and user's (b) Accuracy using f-measure*

4. Conclusion

Machine learning technique approach on satellite imageries perform in statistical Maximum likelihood classification is significantly converging to actual classes. Kappa coefficient derives the significantly different from a random assignment. The study achieved more than 95% classification accuracy in agricultural crops. In future accurate estimation can be project with satellite imageries.

References

Duda, T., and Cantry, M. *Unsupervised Classification of Satellite Imagery: Choosing Good Algorithms.* International Journal of Remote Sensing. 2002. 23 (11) 2193-2212.

Foody, G.M. *Status of Land Cover Classification Accuracy Assessment*. Remote Sensing and Environment. 2002. 80; 185-201.

Gordon, S. *Utilizing Landsat Imagery to Monitor Land Use Change: A Case Study in Ohio*. Remote Sensing of Environment. 1980. 9; 189-196.

Kun Wang, Youchuan Wan, and Shaohng Shen. *Classifications of Remote Sensing Images Using Fuzzy Multi-Classifiers.* IEEE International Conference on Intelligent Computing and Intelligent Systems 2009. (4) 411-414.

Lu, D., and Weng, Q. *A Survey of Image Classification Methods and Techniques for Improving Classification Performanc*e, International Journal of Remote Sensing. 2007. 28 (5) 823-870.

Perumal, K., and Bhaskaran, R. *Supervised Classification Performance of Multispectral Images.* Journal of Computing. 2010. 2 (2) 124-129.

Richards, J.A., 1995: Remote Sensing Digital Image Analysis an Introduction. Springer-Verlag. 265-290.