

Accuracy Assessment of Supervised and Unsupervised Classification using NOAA Data in Andhra Pradesh Region

M. Sreelekha

Master of Technology

Department of Electronics and Communication
Engineering

Sri Venkateswara University College of Engineering,
Tirupati, India.

Dr. S. Narayana Reddy

Professor

Department of Electronics and Communication
Engineering

Sri Venkateswara University College of Engineering
Tirupati, India.

Abstract— The objective of this study is to differentiate NOAA satellite data using NDVI thresholds. Normalized Different Vegetation Index image (NDVI), initially derived from visible and near infrared bands of NOAA satellite. The different areas like vegetation, non-vegetation and water bodies are keenly observed and the thresholds for classifying them are formulated carefully with the help of ground truth information of the study area. The separation of images into different land covers is performed using density slicing. The classification process is completed by color mapping and class labelling. Confusion matrix is used to determine the accuracy of classified image by calculating overall classification accuracy and Kappa coefficient. NDVI based classification is one of best method to classify the NOAA satellite data with a high accuracy.

Keyword —AccuracyAssessment, Normalised different vegetation index (NDVI), Near infrared(NIR), Threshold, Vegetation, Kappa coefficient.

I. INTRODUCTION

Remote sensing data have a wide range of applications, among them land cover mapping has its own significance. The physical condition of the ground surface can be detected using land covers. In the study of land cover dynamics, remote sensing is majorly done by satellites. From last few decades, due to the advancement in technology in remote sensing, there is enhancement in obtaining a large geographical data.

The accurate and timely information about location and spatial configuration as well as growth rates of land covers is provided by sensors. Image classification is process used for land cover mapping from remote sensing data. The major breakthrough in satellite remote sensing is better spatial and spectral resolution and it's been possible by employing advanced sensors. For mapping and monitoring forest/vegetation, both visual and digital analysis techniques have been used. Interpreting various land use classes such as forest, vegetation is done by adapting proper methodology. The analysis of remotely sensed imagery through interpretation is distinguished by following three factors. They are

1. Panoramic overview of remotely sensed imagery.
2. They fall in the region of visible and infrared region of the electro-magnetic spectrum and

3. portraying the Earth's surface at different scales and resolutions.

The necessary spectral and spatial features of the various objects can be obtained through multispectral remote sensing. Technique used for classification of objects is spectral analysis of the radiant energy reflected or emitted by the target. In this paper, the Normalized Difference Vegetation Index (NDVI) values and classifying different land cover types over the selected study area is done by multispectral images. The difference between the sensor spectral radiance of the red band (band4) and the near-infrared band (band5) of satellite image, gives NVDI. Generally, NVDI values are positive for soil and vegetation and theoretically the values of the NDVI vary between -1.0 and $+1.0$.

This paper is divided into 5 sections, section II describes the detailed of Study Area and Data, Methodology used is given in section III. Section IV deals with the obtained experimental results for the proposing approach. Section V contains concluding remarks.

II. DESCRIPTION OF STUDY AREA AND DATA

Andhra Pradesh region displays a lack of continuity of stratigraphic significance which indicates a period of remarkable tranquility. This is an example of Eparchaeon Unconformity [Source: Tirupati Urban Development Authority]. The monsoon bring a moderate rainfall and in summer, temperatures shoot up ranging from 35 to 41 degrees Celsius and in winter the city experience low temperatures ranging 16 to 20 degrees Celsius. The summer season starts from March and may last up to June, followed by rainy season from July and ends in December. The city then experiences winter till the end of February. In this paper, rainy season is considered and the study area experiences maximum rainfall from October to November, which falls under northeast monsoon season.

For the study, images issued by to the Centre of Excellence on "Atmospheric Remote Sensing and Advanced Signal Processing", were collected and used. The images taken are multispectral remote sensing images of Andhra Pradesh region of three different seasons which are listed above .The satellite used here is NOAA, which images the entire region

of earth surface for a period of 16 days .Multispectral band data is provided by the sensors called OLI (Operational Line Imager) and TIRS (Thermal Infrared Sensor) . Table I contains details regarding band designations of NOAA. This study employed the Satellite data over Andhra Pradesh region of 21st January, 30th May and 21st October of 2018 day time, level-1G product, path/row 143/51).

FCC (False Color Composite) images are formed by combining the Visible bands and Near Infrared bands together. The nearest neighbor method is used for resampling images.

The visible and infrared bands of NOAA spectral bands is used to calculate NDVI. For calculation of NDVI technique called per-pixel is been employed. Here the normalized difference between the red band (0.64 - 0.67Pm) and near infrared band (0.85-0.88Pm) is considered for formula mentioned below.

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

Where NIR is the near infrared band value of a pixel and RED is the red band value of the same pixel.

Maximum Likelihood Classification is used for image classification and is carried out for each dataset, training sets were collected from FCC imagery and supervised imagery. Comparison of statistical NDVI values like minimum, maximum, mean and standard deviation is done for various land cover types of different seasons. The same category of land use/cover will have same or similar curves of NDVI and NDVI time series provide seasonal changes. The land-cover change detection and phenological parameter derivations are supported by time-series data of normalized difference vegetation index (NDVI).

The classification of the image based on NDVI values is shown in the Figure1 and the statistics of NDVI are estimated as:

The statistics show images with low NDVI value are usually water bodies (ranging from -0.0175 to -0.328), Built up (ranging from -0.019 to 0.060) and bare soil (ranging from -0.001 to 0.166). Throughout the three seasons, the NDVI value of the dense vegetation is ranging from 0.500 to 0.575, that almost remind consistent, which indicates forest region in Eastern Ghats. On the other hand, for three different seasons, the vegetation in and around the city varies from 0.244 to 0.44.

In this paper, the classification of the study area on the basis of NDVI by using multi season and multispectral satellite data. The dense and thick forest areas are been observed in the Andhra Pradesh region, which is part of eastern Ghat division in geographical division of Indian mountain ranges. A plenty of waterbodies are situated at eastern part of Andhra Pradesh. Vegetation land is extended over east and west regions. Central part is mostly found with built up area and the settlements. The result obtained from study state that there is no large significant NDVI changes are found in dense vegetation and its dynamics, throughout the three seasons. Same methodology can be implemented for retrieving of non-spatial parameters which can affect the vegetation through climatic changes over the study area. The Land Surface

Emissivity (LSE) and Land Surface Temperature (LST) can be estimated from NDVI and the correlation between them can be calculated.

III. MATERIALS AND METHODS

The flowchart of the proposed work to get classified NDVI image was shown in Figure 1. This technique is capable only to process NOAA data. In this study, Channel 1 and Channel 2 are used to calculate NDVI.

Table 1: The NOAA band characteristics shown in Table

Band no	Band designation	Wavelength (µm)	Resolution at nadir
1	Visible red light	0.58-0.68	1.1Km
2	Near Infrared	0.725-1.1	1.1Km
3	3A,Near Infrared	1.58-1.64	1.1Km
	3B Infrared	3.55-3.95	1.1Km
4	Thermal Infrared	10.3-11.3	1.1Km
5	Thermal Infrared	11.5-12.5	1.1Km

Step1:

Before analysis, the images were radiometrically and geometrically corrected. During geometric correction, control points are detected on the topographic maps and the satellite images with RMS errors that are less than two pixels. After that, the images of 2016 and 2018 were registered. A subset image was created from each NOAA image for subsequent analysis.

Step2:

Normalized Difference Vegetation Index (NDVI) was used to identify different land cover types of the study area by using the equation (1).

NDVI was calculated for the year 2016 and 2018 of NOAA data. The results were also analyzed, which shows significant changes in land cover over a period of time in the study area. NDVI ranges between -1.0 to +1.0. The cloud, water, and snow reflect in the visible channel than the near-infrared channel, they take negative NDVI values. Rock, bare soil and man-made structures have an NDVI values are around zero. Vegetation, on the other hand, has strong reflectance in near-infrared there by providing NDVI values close to +1.

Step3:

In the present work, NOAA data of both dates were classified independently based on the NDVI values range from -1 to +1. Based on this NDVI values, the two dates NDVI images were classified into four classes by using NDVI threshold ranges. The sparse vegetation and dense vegetation fall between 0.2 to 1. Similarly, water bodies and areas without vegetation falls between -1 to 0.2. Thereafter, region of interest was selected and supervised classifier of maximum likelihood classification algorithm was used to produce them

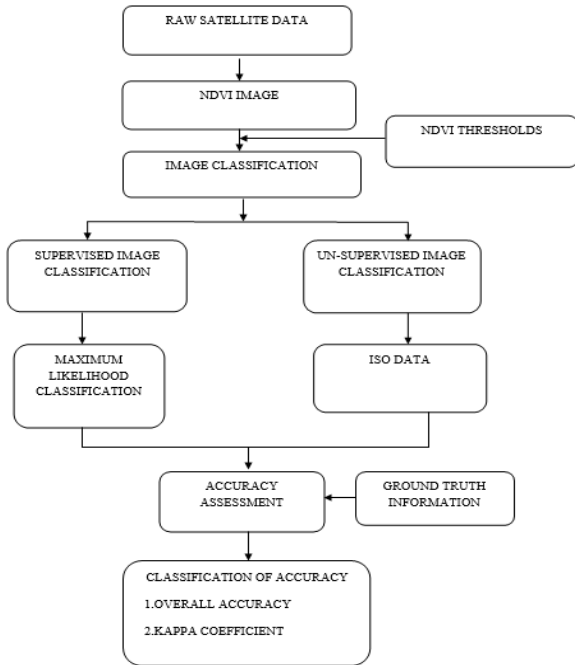


Figure 1:Block Diagram of Accuracy assessment of supervised and unsupervised Classification Using NOAA data.

A. CLASSIFICATION

The image enhancement is initial process done by using decorrelation stretch to enhance the image for more effective visualization, before the classification. The flow chart in figure 1 represents the classification and accuracy assessment process. The experiment further proceeded by calculating the NDVI thresholds. The different combination bands of 5-4-3 (NIR, Red, and Green) are constructed to red, green and blue (RGB color) are been employed here. The RGB image is a standard color for/of infrared (CIR) image.

$$NDVI = \frac{Band5 - Band4}{Band5 + Band4} \quad (2)$$

The experiment is carried out further by locating the vegetation by calculating the threshold of NDVI image. Based on the values of NDVI threshold are shown in

Table 2: Values Of NDVI

SNO	Description	Value of NDVI
1	Very low NDVI	<-0.1
2	Moderate NDVI	-0.1 to 0.25
3	High NDVI	>0.25

The training data classified mainly into three major classes: Water, Non-vegetation and Vegetation area. The results have been kindly observed and analyzed. The NVDI is set as modified gray level image by the process called density slicing. The next consecutive experimental step involves the color mapping and labeling of the satellite image for three classes. The process of classification based on NDVI threshold is thus concluded and the study further involves comparison of the classified image with the ground truth.

B. ACCURACY ASSESSMENT(CONFUSION MATRIX) :

The classified image obtained by experiment is now compared with the ground truth data for analysis of information about accuracy. Both user and producer accuracy were measured to calculate overall classification accuracy. when individual class accuracy divided by the sum of correctly classified pixels, the Producer’s accuracy. Both the misclassified pixels and errors in them where classified into another class and are recorded. On the other hand the user’s accuracy is a calculated when individual class acquired from the classified pixels in same group. The confusion matrix is formulated to calculated the over all accuracy which is obtained for user’s accuracy and producer’s accuracy. The details of overall accuracy is shown below which is acquired by dividing.

$$\text{Overall Accuracy} = \frac{\text{Total no.of correct classified}}{\text{Total no.of pixels}} * 100 \quad (3)$$

The measurement to measure the training pixels with the ground truth data, is by using technique of Kappa coefficient. The Kappa values are in range of +1.0 to -1.0, if the value is the positive value, the it shows high accuracy. A value of zero in Kappa coefficient indicates no correlation in the classification

$$K = \frac{n \sum_{i=1}^p x_{ii} - \sum_{i=1}^p x_{i0} x_{0i}}{n^2 - \sum_{i=1}^p x_{i0} x_{0i}} \quad (4)$$

Where:

- n = total number of pixels
- p = total number of classes
- $\sum x_{ii}$ = total number of elements in confusion matrix
- $\sum x_{i0}$ = sum of row i
- $\sum x_{0i}$ = sum of column i

The results of classification based on the accuracy assessment were obtained and recorded in the next section.

IV. RESULTS AND DISCUSSION

Accuracy of any classification process is to be quantitatively determined and thus the results to be obtained from satellite data. NDVI image was compared with the ground truth information with the pixel that already been categorized. NDVI ratio is used for evaluation of the supervised NDVI threshold for Classification of an area (349 pixels 329 pixels) of image. The grayscale image of NDVI is shown in Figure 2. The color mapping and labelling of three classes is done for the classification process. This is showed and observed in Figure 4. The images are divided into three categories based on NDVI value. The lower coverage is obtained for water class. The medium value of NVDI is obtained for the other class like for vegetation (contains urban, bare land, etc.) The non-vegetation have the highest NDVI. The NDVI image based on threshold values are calculated and the pixels of image are classified, under this supervised classification. Those images whose pixels have NDVI ratio greater than 0.35 (contain moderate and high NDVI values) are designated to the vegetation class.

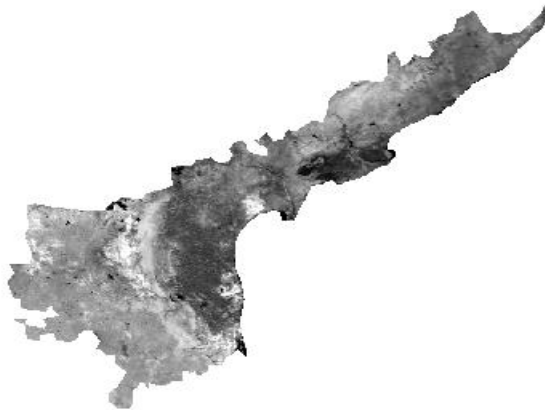


Figure 2: 2016 NDVI image

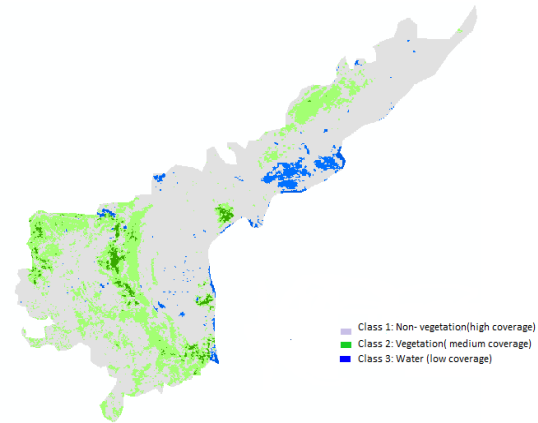


Figure 5: 2018 Classified image

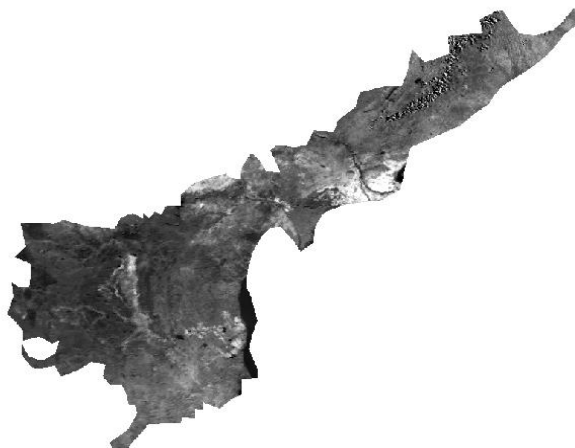


Figure 3: 2018 NDVI image

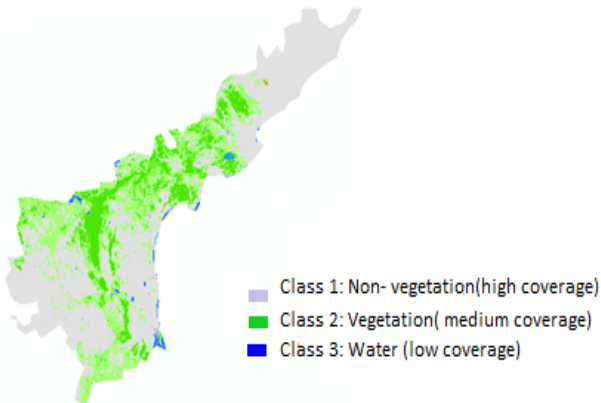


Figure 4: 2016 Classified image

Table 2. is formulated and assessment of the related confusion matrices are done, along with the classification results (overall accuracy and kappa coefficient). From the results, the total pixels are correctly categorized as very high rather than misclassified pixels for every category class.

Table 2: Confusion matrices

	water	Vegetation	Non-vegetation	sum	User's accuracy
Water	25,550	446	2,880	28,876	0.87
Vegetation	0	63,089	778	63,867	0.96
Non-vegetation	65	927	21,024	22,016	0.92
Sum	25,656	64,465	24,686	114,807	
Producer's accuracy	1.00	0.96	0.82	2.78	

For the water class which are into a non -vegetation class, the misclassified pixels are 65 only the misclassified pixels of other classes are also formulated and concluded. The vegetation class are in the other two classes; water and non-vegetation, 446 and 927, respectively. The misclassified pixels for non-vegetation class are also into the other two classes namely water and vegetation are 2880 and 778, respectively. From every row and every column of the confuse on matrix table, one can obtain the user's accuracy and producer's accuracy. The classified pixels of every row and column are divided with the sum of total pixels for every row and column. The result is highly accurate and the overall accuracy is obtained 92.55%. The value of kappa coefficient is positive and in the range of 0.906. The observed results are with high accuracy for obtained kappa coefficient.

The classification process here analyzed by considering the confusion matrix must be quantitatively done to obtain their accuracy of classification for satellite data taken from NOAA. As already mentioned earlier in the supervised classification, the NDVI images are categorized into certain classes based on their NDVI value. The user will have flexibility of choosing the training sample for classification based on one's knowledge.

V. CONCLUSION

NDVI thresholds are used for classification of NOAA satellite data in the study carried out. There are namely three classes of interest which are vegetation, non-vegetation and water. The accuracy of the classification is carried out by calculating the overall classification accuracy and Kappa coefficient. The results make the fact clear that NDVI - based method is capable for the classification of NOAA satellite data with a high and promising accuracy.

ACKNOWLEDGMENT

The authors are extremely grateful to the Centre of Excellence on "Atmospheric Remote Sensing and Advanced Signal Processing", at Department of ECE, SV University College of Engineering, Tirupati, Andhra Pradesh, India, for providing necessary resources to carry out the present work.

REFERENCES

- [1] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation," 1973.
- [2] D. Xu, "Compare NDVI Extracted from Landsat 8 Imagery with that from Landsat 7 Imagery," *Am. J. Remote Sens.*, vol. 2, no. 2, p. 10, 2014.
- [3] A. Ahmad, B. M. Aboobaider, M. S. M. Isa, N. M. Hashim, M. Rosul, S. Muhamad, and S. Man, "Temporal changes in urban green space based on normalized difference vegetation index," *Appl. Math. Sci.*, vol. 8, no. 55, pp. 2743–2751, 2014.
- [4] R. Geerken, B. Zaitchik, and J. P. Evans, "Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity," *Int. J. Remote Sens.*, vol. 26, no. 24, pp. 5535–5554, 2005.
- [5] a. K. Bhandari, a. Kumar, and G. K. Singh, "Feature Extraction using Normalized Difference Vegetation Index (NDVI): A Case Study of Jabalpur City," *Procedia Technol.*, vol. 6, pp. 612–621, 2012.
- [6] H. Fan, X. Fu, Z. Zhang, and Q. Wu, "Phenology-Based Vegetation Index Differencing for Mapping of Rubber Plantations Using Landsat OLI Data," *Remote Sens.*, vol. 7, no. 5, pp. 6041–6058, 2015.
- [7] B. Satyanarayana, K. A. Mohamad, I. F. Idris, M.-L. Husain, and F. Dahdouh-Guebas, "Assessment of mangrove vegetation based on remote sensing and ground-truth measurements at Tumpat, Kelantan Delta, East Coast of Peninsular Malaysia," *Int. J. Remote Sens.*, vol. 32, no. 6, pp. 1635–1650, 2011.
- [8] A. Viña, A. a. Gitelson, A. L. Nguy-Robertson, and Y. Peng, "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3468–3478, 2011.
- [9] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, 1991. NDVI, the Foundation for Remote Sensing Phenology