Impact of features on classification accuracy of IRS LISS-III images using artificial neural network

Mr.Anand Upadhyay, Dr (Mr.) Santosh Kumar Singh , Mr.Vipin.G.Shukla

Department of IT Thakur College of Science &Commerce ,Thakur Village, Kandivali (E), Mumbai-400101, India

ABSTRACT

The proposed research paper uses the artificial neural network with back propagation of error is used to classify the LISS-III satellite images. It is a multi-spectral images which are stacked together and there different feature are used for the classification. In this classification the comparative study is performed on two different types of training sets which are used for the classification training. The First training set consist of the R, G, B colour value of the image in this case accuracy and kappa coefficient was 95% and 0.9331 respectively. Second training set consist of R, G, B, mean, mode, median, variance and standard deviation of colour feature and in this case the accuracy and kappa coefficient was 97% and 0.9652 respectively. So after this study and comparative analysis it is observed and concluded that there is the impact of features on the classification. If the number features are less then there is fall in the classification accuracy but increase in the classification accuracy with respect to increase in the number of features in the training sets.

Keywords:- LISS-III satellite Image, mean, mode, median, variance, standard deviation, neural network, and classification.

1.INTRODUCTION

Classification is one of the data mining methods which are used to classify the object into predefined group. It is the most frequently used decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. The classification also plays very important role in the remote sensing and satellite image classification. There are number of application i.e. encompassing reconnaissance creation of mapping products for military and civil application, evaluation of environmental damage, monitoring of land use and land cover urban planning, growth regulation, soil assessment and crop yield appraisal where image classification plays major role to study and monitor all these things[1]. Due to the advancement in the area of remote sensing, data mining tools, digital image processing and high computational computers there are number of students, academician, research scholars and different organizations are attracted towards the remote sensing and the different application of remote sensing. They are try to find out different ways through which they can use the satellite images for various purposes either related to different resource evaluation, analysis, planning, management or monitoring purpose. So, in all these application one important application of satellite image is land use and land cover classification. Land use and land cover classification plays very important role for the development of thematic map and to study the area covered by different resources. The satellite based land use and land cover classification plays important role to map the different resources which saves the time of doing so much survey and field work. It also covers those area where the survey and field work cannot easily perform or difficult to perform. The land use and land covered classified images are also used to study the change detection which are carried out by the human or the nature. We can also study the impact of human settlement and its impact on the natural resources. There are number of classification techniques are available i.e. KNN (K-nearest neighbour), Decision tree, ANN (artificial neural network), SVM (Support vector machine), Bayesian classifier and Discriminate base classification. All these classification techniques are basically classified in two categories which are known as supervised and unsupervised classification. An unsupervised classification technique, in which the objects are classified based on its similarity which is called as cluster. On the other hand in supervised classification technique classification performed based on the training set which is called as the learning from the example. Here, in this paper artificial neural network is design to classify the IRS-P6 LISS-III satellite image for various land use and land cover classification. The artificial neural network with back propagation of error applies the supervised classification which requires the training set of data. So, here in the paper the artificial neural network based comparative analysis is performed and it is tried to find out what is the impact of feature in the case of supervised classification. In the comparative study of ANN based classification there are two types of samples are used in which one samples has three feature where as in the second one there are eight features are used for the training purposes. So, with comparative analysis of ANN based classification, it is found that there is the impact of attribute on the classification because the

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 10, October 2014   **ISSN 2319 - 4847**

accuracy of classification of neural network which trained with the three features vector was less compare to the accuracy of classification where the training was performed with help of eight different features.

A. Survey and Field Study

The survey and field study plays very important role in case of satellite image classification. In the survey and field study the area or region which we want to classify that survived and the analyst tries to locate these areas on the LISS-III data. These areas are known as "training sites" [2]. These data are used to guide or train the classifier and then classifier is able to classify the unknown data based on the train data sets. Here, the data is collected by maps, and personal experience.

B.Study area and characteristics

The Resourcesat - 1 is designed to provide multispectral, monoscopic and stereoscopic imageries of the earth's surface with its advanced on-board sensors. Indian Remote Sensing (IRS-P6) satellite gives LISS-III (Linear Imaging and Self Scanning Sensor) data. Classified land use and land cover map generated from LISS-III data is used to assess the land cover by human settlement, water, forest and mangroves. The data set of Mumbai region acquired for this research was collected via IRS-P6 resourcesat-1 satellites using LISS-III sensors in the multispectral mode by NRSA, Hyderabad, Andhra Pradesh (A.P), and India. The LISS-III is a multi-spectral camera operating in four spectral bands. The LISS-III data sets consist of four different bands which are 0.52-0.59 microns (B2), 0.62-0.68 microns (B3), 0.77-0.86 microns (B4) and 1.55-1.70 microns (B5) of images. There are three bands in the visible and one near infrared. LISS-III provides data with a spatial resolution of 23.5 m unlike in IRS-P6 (where the spatial resolution is 70.5 m).The LISS-III data sets are widely used for Land use/ Land cover, Urban planning, biodiversity characterization, Forest survey, wet land mapping, environmental impact, crop acreage and production estimation of major crops, drought monitoring and assessment based on vegetation condition, snowmelt run off estimation and so on [3].

2. ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is biologically inspired from the human biological neural network which is called as "Neurons". The artificial neural network often called as "neural network". The artificial neural network is a mathematical model or computational model based on biological neural networks. The artificial neural network is very appropriate and suitable where there is no mathematical model is applicable between the attribute and target. The artificial neural network consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. The artificial neural network is dynamic in the nature because in most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

A. Feed forward neural network with back propagation of error

The artificial neural network is very popular and widely used for classification purposes. The feed forward neural network is very simple and popular neural network. The feed forward neural networks consist of basically three parts which are input layers, hidden layers and output layers. In feed forward neural network the information moves in only one direction which is in the forward direction. The forward means information follows from the input nodes to the hidden nodes and from hidden nodes to the output nodes. There are no cycles or loops in the network. The data processing can extend over multiple (layers of) units.
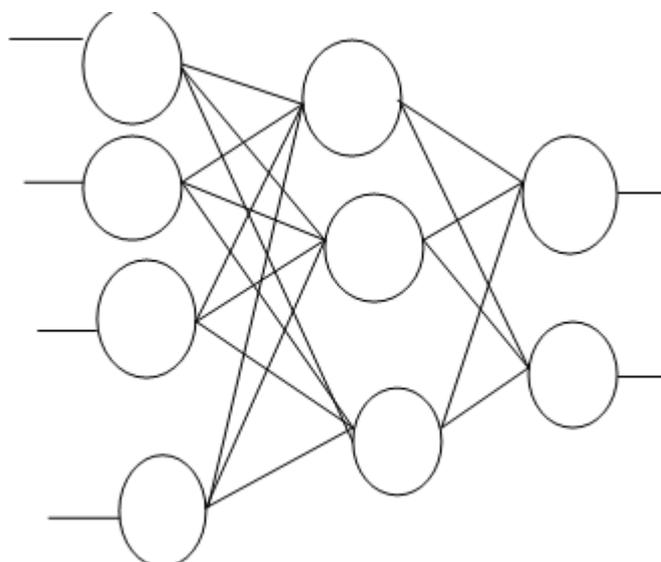

Fig.1. Architecture of artificial neural network

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 10, October 2014   ISSN 2319 - 4847

B.Feed forward neural network using back propagation of error algorithm
- Decide input, target and testing data.
- Initialize the weight and bias.
- Calculate the feed forward Neural Network output.

Y=b+∑x (i)*w (i);
- Match the output with target.
- Calculate the error= (Oi -Ei)
- Update all the weight and bias of the Neural Network.

W (new) =W (old) +ΔW;

b (new) =b(old)+Δb;
- Repeat the steps until the error will not reduced.

C.   Training of artificial neural network
The training of artificial neural network is basically classified into two categories which are as follows.

1) Supervised Learning
The supervised learning in artificial neural network is used for the purpose of classification. In supervised learning initially the neural network is trained with the known data sets. This data sets consist of the different features and there labelled classes. Once the artificial neural network is trained then it is ready to classify the unknown data set. So, when the unknown data set is provided then based on its past learning it will allocated the classes. This type of learning is called as supervised learning.

2) Unsupervised Learning
The unsupervised learning is very much opposite to the supervised learning. The unsupervised learning is basically used for the purpose of clustering in the data sets. In unsupervised learning the neural network do not has the training data set here based on the similarity that is mean distances among the feature the artificial neural network divide the data into different region which is called as clusters. So, this type of learning is called as unsupervised learning.

3.METHODOLOGY

The algorithm is developed in MATLAB R2010. It uses a feed forward algorithm. LISS-III satellite image consist of four different bands of images. Initially, before going to extract the feature of satellite image, all four different bands are stacked together to get the one RGB image. The following flow chart represents the steps involved in satellite image classification.
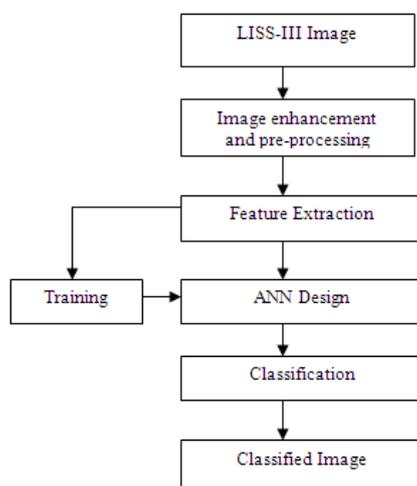

Fig.2. Flow chart for Classification

A.   Feature Extraction
In order to classify the satellite images, the first step is feature extraction. As paper is based on the impact of features on the classification here two types of training data samples are prepared. The first training sample data set consist of the R (Red), G (Green), B (Blue) colour features of each pixels. The second training sample data set consist of R(Red),G(Green),B(Blue) colour features and simultaneously it consist of mean, mode, median, variance and standard deviation of colour features of each pixels. Based on the features of known classes the neural network is trained. During testing when the features are provided for pixels which are not having the classes then it will give the class based on their learning.

B.Artificial Neural Network Design
The Matlab is very powerful mathematical computational model which provides the tool for neural network based classification. The feed forward back propagation of neural network is designed by using the neural network toolbox.

The data sets are pre-processed as the Matlab toolbox required. The neural network object in the Matlab toolbox expects the samples along columns and its features along rows. Our dataset has its samples along rows and its features along columns. After pre-processing of sample data the next step is to create a neural network (feed forward back propagation network) that will learn to identify the classes. Since the neural network starts with random initial weights, the results will differ slightly every time it is run. The random seed is set to avoid this randomness. A 1-hidden layer feed forward network is created with 30 neurons in the hidden layer. After all these setting the network is ready to be trained. The both types of training data sets are used to train the neural network at a time and used to test on the testing data separately. Each samples data set is automatically divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy.

C. Classification
The neural networks which are trained differently with two different training data sets are used to classify the desired image. The trained neural network is ready to classify the desired image. The testing can be done with a separate testing set which is created while creating training set and based on that the classification accuracy calculated using confusion matrix and Kappa coefficients.

4.DATA ACQUIRED

LISS-III data set of Indian region is freely available online. LISS-III data sets are available in four different bands. The data is acquired in April 2013.The field work is carried out in the month of July and september2013.

5.RESULT

The Neural Network trained with sample data. After training, the testing is done using testing dataset and the accuracy assessment is done through the confusion matrix and Kappa co-efficient. The neural network is used to classify the desired image. Here the two different types of training data sets are used and separately tested on the testing data set and there confusion matrix and Kappa coefficient are calculated. The confusion matrix and kappa coefficient are for the data set which is tested by neural network which is trained with the first data set which consist of only three feature vectors.

TABLE 1.CONFUSION MATRIX OF FIRST DATA SET TRAINED ANN

Class	Water	Forest	Mangroves	Settlements	Total	User's Acc
Water	218	0	0	0	218	100%
Forest	2	184	11	0	197	93.40%
Mangroves	0	9	76	5	90	84.44%
Settlements	0	0	0	71	71	100%
Total	220	193	87	76	547	
Producer's Accuracy	99.09%	95.34%	87.36%	93.42%		

Accuracy= (547/576)*100;
=95.31
=:95%

$$\hat{k} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(X_i + * X + i)}{N2 - \sum_{i=1}^{r}(X i + * X + i)}$$

k = 0.9331(Very Good)

The confusion matrix and kappa coefficient are for the data set which is tested by neural network which is trained with the first data set which consist of only three feature vectors.

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 10, October 2014   **ISSN 2319 - 4847**

TABLE 2. CONFUSION MATRIX OF SECOND DATA SET TRAINED ANN

Class	Water	Forest	Mangroves	Settlements	Total	User's Acc
Water	220	2	1	0	223	98.65%
Forest	0	187	9	1	197	94.92%
Mangroves	0	4	75	0	79	94.94%
Settlements	0	0	2	75	77	97.40%
Total	220	193	87	76	557	
Producer's Accuracy	100%	96.89%	86.21%	98.68%		

Accuracy= (557/576)*100;

=96.70

=:97%

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(X_{i+} * X_{+i})}{N2 - \sum_{i=1}^{r}(X_{i+} * X_{+i})}$$

k = 0.9652(Very Good
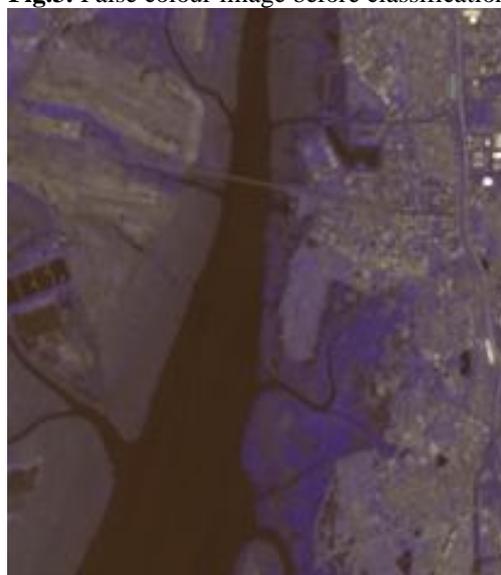

Fig.3. False colour image before classification
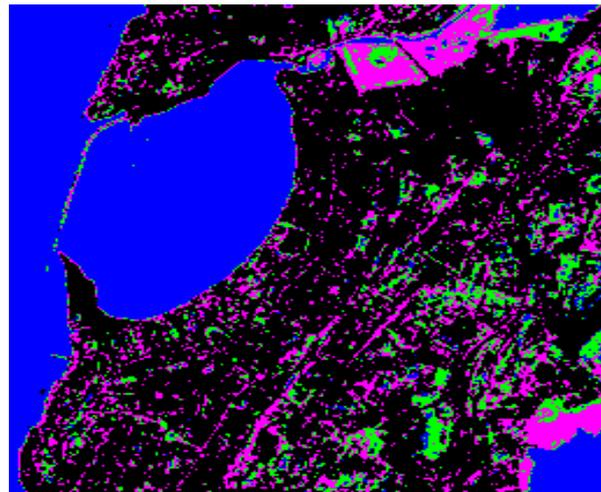

Fig.4. False colour image before classification

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 10, October 2014   **ISSN 2319 - 4847**

Fig.5. Colour image after the classification

Fig.6. Colour image after the classification

water   Forest   Mangrove   settlement
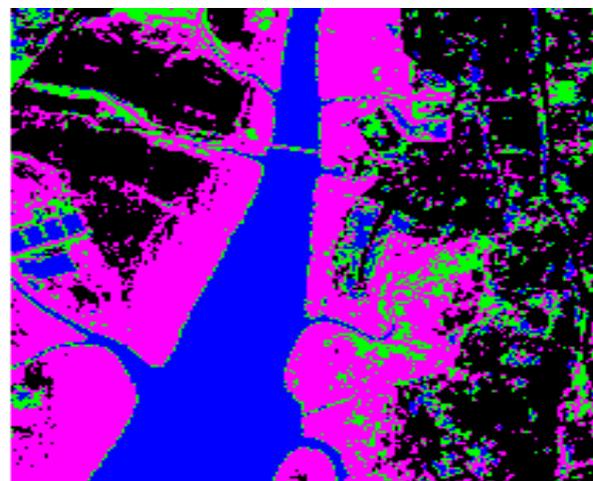
6. CONCLUSION

In this paper back propagation neural network has been used for classifying LISS-III satellite image and studying the impact of features on the image classification. Here, from the above confusion matrix and kappa value calculated for the both types of data sets it found that the accuracy of classifying the image with less feature (First training set) accuracy 95%and Kappa coefficient 0.9331 is less the accuracy classifying the image using more feature (Second training set) accuracy 97% and Kappa coefficient 0.9652 is more than the first one. So, it is concluded that there is the impact of number of feature on the image classification.

7. FUTURE ENHANCEMENT

The Impacts of feature is tested and comparative study is performed using the artificial neural network. The above impact of feature on classification can also tested with the other classifiers which Decision tree, Support vector machine and Bayesian classifier and results are compared with the artificial neural network.

ACKNOWLEDGMENT

REFERENCES

[1] K Perumal and Bhaskaran, Supervised Classification Performance of Multispectral images, Journal of Computing, Volume 2, February 2010, ISSN 2151-9617.

[2] V.Turkar, 2010, Polarimetric SAR image classification by using artificial neural network, ICWET, Organized by TCET, Kandivali (E), Mumbai, India, PP.48-52, and ISBN: 978-1-60558-812-4.

[3] Madhubala.M, S.K.Mohan Rao and Ravindra G Babu, 2010, Classification of IRS LISS- III Images by using Artificial Neural Networks, IJCA, Special Issue on RTIPPR (3):152–157.

[4] Aria, E.Hosseini, J.Amini and M.R.Saradjian, Back Propagation Neural Network for Classification of IRS-1D Satellite Images,Tehran University, Iran. Print.

[5] Hagan, M. T., and M. Menhaj, (1994), Training feed forward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, vol. 5, no. 6, pp. 989-993.

[6] Dr. M.Seetha, G. Malini Devi and Dr. K.V.N.Sunitha, Performance Assessment of Feature Based Image Retrieval using Neural Networks ,Volume 2, Issue ICTM 2011, February 2012, ISSN Online: 2277-2677.

[7] K. Saha, M. K. Arora, E. Csaplovics and R. P. Gupta, Land Cover Classification Using IRS LISS III Image and DEM in a Rugged Terrain: A Case Study in Himalayas, Geocarto International, Vol. 20, No. 2, June 2005.

[8] Pakorn Watanachaturaporn, Manoj K. Arora, and Pramod K.Varshney, Multi source Classification Using Support Vector Machines: An Empirical Comparison with Decision Tree and Neural Network Classifiers, American Society for Photogrammetry and Remote Sensing, Vol. 74, No. 2, February 2008, pp. 239–246.

[9] M. Seetha, K. V. N. Sunitha, and G. Malini Devi, Performance Assessment of Neural Network and K-Nearest Neighbour Classification with Random Sub windows, International Journal of Machine Learning and Computing, Vol. 2, No. 6, December 2012.

[10] H.R. Keshtkara, H. Azarnivand, H. Arzani, S.K. Alavipanah, F. Mellati, Land Cover Classification Using IRS-1D Data and a Decision Tree Classifier , Keshtkar et al. / DESERT 17 (2013) 137-146.

[11] Giorgos Mountrakis, Jungho Im, Caesar Ogole, 2010, Support vector machines in remote sensing: A review, ISPRS Journal of Photogrammetry and Remote Sensing.

[12] P. Kanungo and S. Sarkar, Use of Multi-Source Data Sets for Land Use/Land Cover Classification in a Hilly Terrain for Landslide Study, Disaster & Development, Vol. 5 No. 1 April 2011:35-51.