1. Project Title : Applications of data mining techniques in improving the classification accuracy of high resolution satellite data

2. Scope and Objectives:

The commonly used automatic methods of satellite image classification, supervised or unsupervised classification algorithms, are good enough when used with low to moderate resolution remotely sensed data. But, in the case of high resolution images like IRS P6 LISS IV, IKONOS or CARTOSAT, showing a diversity of land cover forms, the traditional algorithms have not been found adequate to obtain satisfactory results, because of high variance in the data. In high resolution images, spatial neighbouring pixels are highly correlated and in the most instances, they belong to the same land cover class. Even, most commonly used Maximum Likelihood classifier fails when it works with high resolution data. The proposed study is to investigate the applicability of data mining based clustering technique to improve the classification accuracy of high resolution satellite images which will not only involve the classification of single pixel, but of entire objects, into which the information content of the satellite image is distributed. Following are the objectives identified for the study-

I. To provide a comprehensive review of different clustering techniques used for high resolution image classification.
II. To try the existing data mining applications for satellite image classification.
III. To develop a new spatial data mining technique (clustering) to improve the classification accuracy of high resolution data.

3. Centre/Unit : North Eastern Space applications Centre (NESAC)

4. Collaborating Agency : Tezpur University, Assam

6. Funding Agency : EOAM/DOS

7. Study Are: The study area covers Barapani, Meghalaya and a part of Jorhat district, Assam.

8. Methodology : We proposed to utilise data mining techniques such as neural networks, Fuzzy sets, rule-based systems, machine learning etc. either alone or in combination, to analyze and extract patterns from the data. In this project, we reviewed the existing classification algorithms available for satellite image classification. The project research is proposed to investigate new technologies to enable us to   retrieve and discover (mine) information on high resolution satellite images in order to develop a technique (classification) with these capabilities to solve the problem of high variance data analysis.

9. Data Used : Cartosat-I, IRSP6 LISS IV, Landsat TM, IKONOS etc.

10. Status of the Project: Completed

11. Results:
The research work carried out under this study has been completed. The proposed algorithms were tested on IKONOS, QUICKBIRD, CARTOSAT-I, IRSP6L4 and LANDSAT TM satellite imageries. An approach for extraction of linear features from high resolution Indian satellite imageries has been developed. Under this study, an effective fuzzy-NN (nearest neighbor) approach for classification of Landsat TM has been developed. It basically aims to improve the classification performance of the conventional NN based classifiers by appropriate use of fuzzy membership functions on a set of object feature derived from the image objects. This approach has been tested on IRSP6L4 multispectral imageries and results were found better than its other counterparts such as Maximum likelihood (MLC), Support Vector Machine (SVM) and Decision Tree (DT) classifiers in terms of accuracy (Kappa Index Analysis: KIA) and capability of handling mixed pixel problem.

a. CARTOSAT-1 data	b. Case-1, L1,H	c. Case-1, L2,H

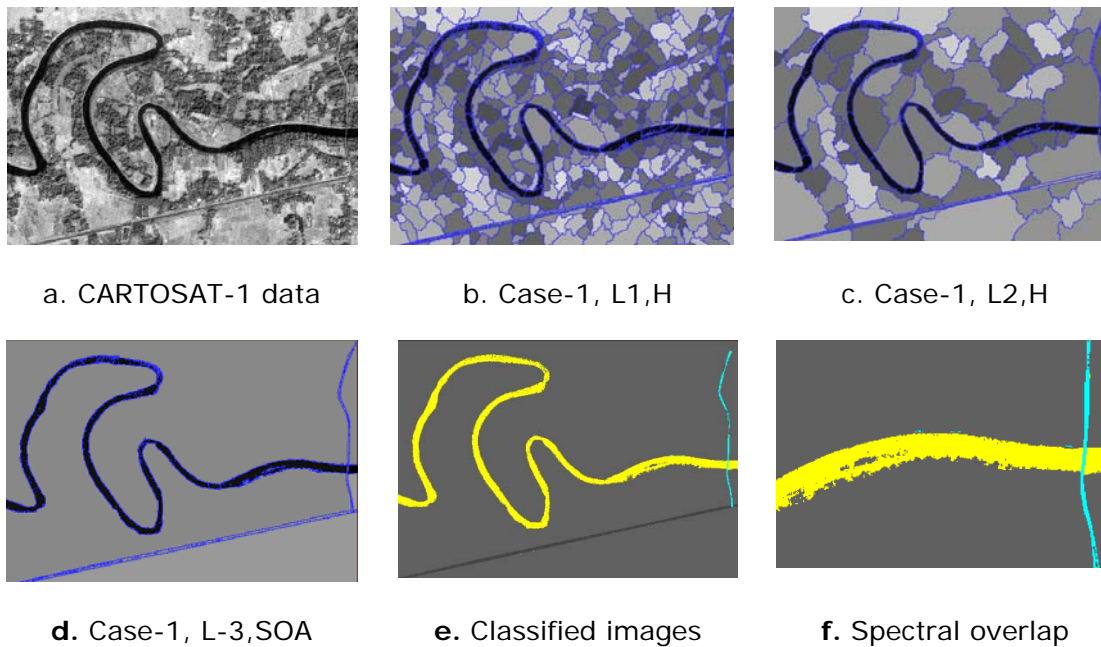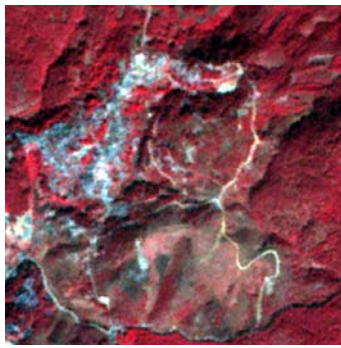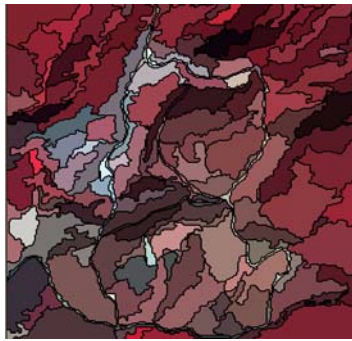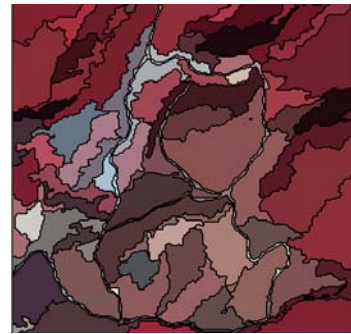| **d.** Case-1, L-3,SOA	**e.** Classified images	**f.** Spectral overlap

Fig.1 Image segmentation and classification for CARTOSAT-1 images
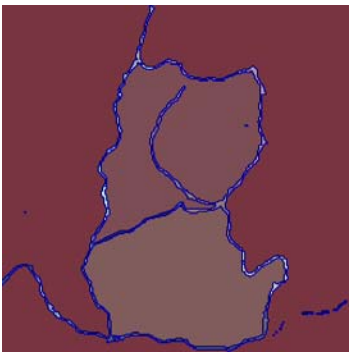(Yellow line indicates river, cyan roads and black line indicates railway line)
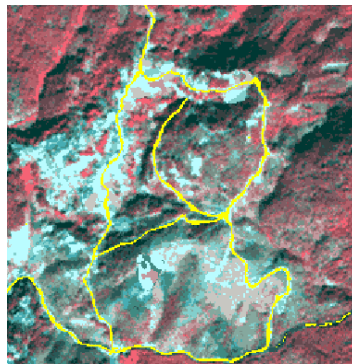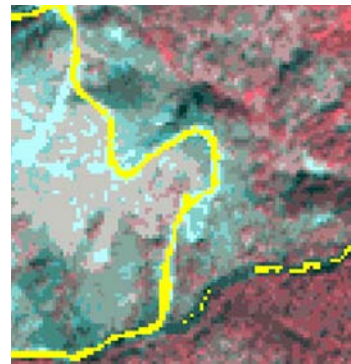
| a. IRS P6 LISS IV MX data | b. Case-3, L1,H | c. Case-3, L2,H |

| d. Case-3, L-3,SOA | e. Classification of roads | f. Failed to classify |

Fig.2 Image segmentation and classification for IRS P6 L4 MX images

12. Utilization/Success Stories:

Some of the significant research findings achieved under this study have been attempted to be published in the reputed journal, book etc. and accepted for publication.

13. Duration: 3 years (2007-2010)