Project Example

Multi Resolution Spatial Data Mining for Assessing Land use Patterns

Uttam Kumar[1], Chiranjit Mukhopadhyay[2], Ramachandra T. V.[3*]

1. *Energy Research Group, Center for Ecological Sciences; Department of Management Studies; & Centre for Sustainable Technologies, Indian Institute of Science, Bangalore, India Email: uttam@ces.iisc.ernet.in*
2. *Department of Management Studies, Indian Institute of Science, Bangalore, India Email: cm@mgmt.iisc.ernet.in*
3. *Energy Research Group, Centre for Ecological Sciences; Centre for Sustainable Technologies; & Centre for Infrastructure, Sustainable Transport and Urban Planning (CiSTUP), Indian Institute of Science, Bangalore, India, Email: cestvr@ces.iisc.ernet.in*

Abstract: Spatial Data Mining or Knowledge Discovery in Spatial Databases (KDSD), i.e. mining knowledge from large amount of spatial data, acquired at various time intervals and from different sensors, is a highly demanding field. Spatial data mining helps in unraveling interesting and previously unknown, but potentially useful patterns from spatial datasets which is more challenging because of the diversity of multi sensor's multi resolutions (spatial, spectral and temporal) data types with spatial relationships, autocorrelation, etc.

***Address for Correspondence:**
Dr. T.V. Ramachandra
Energy Research Group,
Centre for Ecological Sciences,
New Biological Sciences Building,
Indian Institute of Science,
Bangalore – 560 012,
INDIA
Tel : 91-80-22933099
Fax : 91-80-23601428 [CES-TVR]
E-mail:cestvr@ces.iisc.ernet.in
WebURL: http://wgbis.ces.iisc.ernet.in/energy
http://wgbis.ces.iisc.ernet.in/foss

Spatial data acquired from multi resolution remote sensing sensors (in raster format) have been used to provide insights to landscape dynamics through comprehensive understanding of land use and land cover (LULC) pattern of a region. Detailed and accurate inventorying, mapping and monitoring of LULC at a local/regional scale have been possible with the availability of various medium to high spatial resolution sensors (such as Landsat, IRS LISS-III/IV, SPOT, IKONOS, etc.). The land use (LU) information are derived using image processing techniques based on the spectral properties of objects in the bands to assign them into a user defined class label.

Successful classification of spatial data into thematic information requires efficient and optimal image processing. Appropriate classification techniques and feature selection would enhance the classification accuracy. In this context, this chapter presents various advanced supervised pattern classification algorithms such as Maximum Likelihood Classifier, Decision Tree, K-Nearest Neighbor, Neural Network, Random Forest, and Contextual Classification using sequential maximum a posteriori estimation, and Support Vector Machine, apart from their implementation on multi-resolution data acquired from different remote sensing satellites along with an assessment of the classifiers' performance. Hybrid Bayesian Classifier a novel classification technique is presented with case studies in the last section of the chapter.

Keywords: Bayesian classifier, classification, machine learning, multi-resolution data, remote sensing, spatial data mining

1. INTRODUCTION

Advances in data acquisition, storage and database technology have generated enormous amount of data, necessitating efficient information extraction, which has led to an emerging area of data mining or KDD (knowledge discovery in databases). Data mining integrates several domains such as machine learning, database management, data visualization, statistics and information theory. Although numerous studies have been carried out on data mining of relational and transaction databases (Agarwal and Srikant, 1994; Fayyad et al., 1996; Han and Fu, 1996; Piatetsky-Shapiro and Frawley, 1991), data mining of spatio-temporal databases, object oriented databases, multimedia databases, etc. still remain largely unexplored. Thus spatial data mining or KDSD (knowledge discovery in spatial databases) is an active field of research. In this chapter, the focus is on spatial data mining i.e., discovery of interesting knowledge from geospatial data. Spatial data obtained from remote sensing (RS) provides land use and land cover (LULC) patterns of the area under imaging that permit frequent updation of maps nearly on a real-time basis.

Diverse LULC features of the Earth's surface based on their inherent spectral reflectance and emittance properties will have different combination of digital numbers (DNs) (pixel's intensity or grey value) in the image. Radiance/reflectance measurements obtained in various wavelength bands for each pixel provide spectral patterns (referred to as the data henceforth) that can be classified and correlated to different LULC classes on the ground using statistical learning. Recent advances in statistical learning theories have generated promising tools and techniques in

the field of pattern recognition applicable to RS data for deriving the information on land use (LU) classes like urban areas, agricultural land, water bodies, etc. (Kwan et al., 1994; Fukushima et al., 1998; Gori et al., 1998; Lee, et al., 2006). Spectral pattern recognition in LULC classification utilizes spectral information of multiple bands of remote sensors as the basis for automated image classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization (Lillesand and Kiefer, 2002). Performance of spectral classification depends on parameters such as definition and representation of classes, selection of appropriate features, classifier design, selection of training and test samples, and evaluation of classification accuracy. The overall idea of per-pixel classification (hard classification) procedure is to categorize all pixels in an image into LULC classes or themes automatically using either unsupervised classification (such as ISODATA clustering) or supervised spectral classification techniques (such as Maximum Likelihood Classifier).

Unsupervised classification techniques do not require training data and form clusters based on inherent properties of the pixels such as spectral distance to class means, etc. K-Means and ISODATA (Iterative Self-Organizing Data) clustering, for example, begin with arbitrary cluster means and each time clustering repeats, the means of these clusters are shifted. The new cluster means are used for the next iteration. The ISODATA utility repeats clustering of the image until either (a) a maximum number of iterations have been performed, or (b) a maximum percentage of unchanged pixels have been reached between two iterations (Tou and Gonzalez, 1974; Jain and Dubes, 1998; PCI Geomatics Corp; Memarsadeghi et al., 2003).

Supervised classifiers need reference class samples (training data, usually obtained from ground) to predict or classify the unknown pixels in the image. The pixel categorization is done by specifying the numerical descriptors of the various LU types and involves three stages – (i) training stage: identifying representative training areas and developing a numerical description of the spectral attributes of each LU class type in the image, known as training set, (ii) classification stage: each pixel in the image dataset is categorized into the LU class that it resembles most closely, and (iii) output stage: the process consists of a matrix of interpreted LU category types (Lillesand and Kiefer, 2002). However, with heterogeneous and highly fragmented complex landscapes, selecting sufficient training samples becomes rather difficult and the problem worsens with the medium or coarse resolution data due to the presence of mixed pixels. Selection of training samples must take into account the spatial resolution of the RS data considering the complexity of landscape along with the appropriate data mining (classification) technique or scheme. In this context, performance evaluation of different classification algorithms on varying resolution data would be desirable as different classification methods have their own genesis and merits. Although many classification approaches have evolved over time, appropriateness of the approach suitable for features of interest in a given study area is not yet fully understood (Lu and Weng, 2007), necessitating qualitative and quantitative evaluation of the performance of classifiers with multi-resolution data.

The objective of this chapter is to discuss various advanced data mining techniques (such as Maximum Likelihood Classifier, Decision Tree, K-Nearest Neighbour, Neural Network, Random Forest, Contextual Classification using sequential maximum a posteriori estimation, and Support Vector Machine) and select the best classifier for varying spatial resolution data such as high resolution-IKONOS (4 m), medium spatial resolution sensor data-IRS LISS-III MS (23.5 m), Landsat TM (30 m) and low resolution data-MODIS (250 m). Understanding the strengths and weaknesses of variants of multi resolution data of RS sensors are essential for the selection of suitable RS image classification method (Lu and Weng, 2007). This requires prior knowledge of factors such as scale, characteristics of the study area, the availability of data and their characteristics, economic viability, time constraints, etc. The application at the user's end must also take the scale and image resolution into account.

The chapter is organized as follows: section 2 discusses the advanced classifiers and section 3 details the data used in the classification experiments. Section 4 (4.1 to 4.4) presents the algorithm's implementation considering different sensor's data and assessment of their comparative performance with changing spatial and spectral resolutions, followed by discussion in section 5. A new Hybrid Bayesian Classifier is presented in section 6 with concluding remarks in section 7.

2. DATA MINING ALGORITHMS

Seven advanced data mining algorithms such as Maximum Likelihood Classifier, Decision Tree, K-Nearest Neighbor, Neural Network, Random Forest, Contextual classification using sequential maximum a posteriori estimation and Support Vector Machine are presented in this section.

At the outset, let the spectral classes in the data be represented by ω_n, $n = 1, \ldots, N$, where N is the total number of classes, and $(\mathbf{x} \mid \omega_n) : N(\mu_n, \Sigma_n)$. Let if $X_1, \ldots, X_{m_n}$ denote the $(m \times 1)$ grey scale values across the M spectral bands of mn sampled pixels (observations or training samples which are independent and identically distributed (i.i.d.) random variables), belonging to the nth spatial class, then

$$\mu_n = \frac{1}{m_n} \Sigma x_i \tag{1}$$

$$\Sigma_n = \frac{1}{m_n} \Sigma (x_i - \mu_n)(x_i - \mu_n)^T \tag{2}$$

2.1 Maximum Likelihood Classifier (MLC)

Bayes' decision theory forms the basis of statistical pattern recognition based on the assumption that the decision problem can be specified in probabilistic terms (Wölfel and Ekenel, 2005). MLC considers variance and covariance of the category's spectral response pattern (Lillesand and Kiefer, 2002) through the mean vector

and the covariance matrix based on the assumption that data points follow Gaussian distribution. The statistical probability of a given pixel value being a member of a particular class is computed and the pixel is assigned to the most likely class (highest probability value).

$p(\omega_n \mid \mathbf{x})$ gives the probability that the pixel with observed column vector of DNs (digital numbers) $\mathbf{x}$, belongs to class ω_n. It describes the pixel as a point in multispectral (MS) space (M-dimensional space, where M is the number of spectral bands). The maximum likelihood (ML) parameters are estimated from representative i.i.d. samples. Classification is performed according to

$$\mathbf{x} \in \omega_n \text{ if } p(\omega_n \mid \mathbf{x}) > p(\omega_j \mid \mathbf{x}) \; \forall \, j \neq n \tag{3}$$

i.e., the pixel vector $\mathbf{x}$ belongs to class ω_n if $p(\omega_n \mid \mathbf{x})$ is largest. The ML decision rule is based on a normalized estimate of the probability density function (p.d.f.) of each class. MLC uses Bayes decision theory where the discriminant function, $g_{\ln}(\mathbf{x})$ for ω_n is expressed as

$$g_{\ln}(\mathbf{x}) = p(\mathbf{x} \mid \omega_n)\, p(\omega_n) \tag{4}$$

where $p(\omega_n)$ is the prior probability of ω_n, $p(\mathbf{x} \mid \omega_n)$ is the p.d.f. (assumed to have a Gaussian distribution for each class ω_n) for pixel vector $\mathbf{x}$ conditioned on ω_n (Zheng et al., 2005). Pixel vector $\mathbf{x}$ is assigned to the class for which $g_{\ln}(\mathbf{x})$ is greatest. In an operational context, the logarithm form of (4) is used, and after the constants are eliminated, the discriminant function for ω_n is stated as

$$g_{\ln}(\mathbf{x}) = (\mathbf{x} - \mu_n)^T \, \Sigma_n^{-1} \, (\mathbf{x} - \mu_n) + \ln |\Sigma_n| - 2 \ln p(\omega_n) \tag{5}$$

where Σ_n is the variance-covariance matrix of ω_n, μ_n is the mean vector of ω_n. Equation (5) is a special case of the general linear discriminant function in multivariate statistics (Johnson and Wichern, 2005) and used in this current form in the RS digital image processing community. A pixel is assigned to the class with the lowest $g_{\ln}(\mathbf{x})$ in equation (5) (Duda et al., 2000; Zheng et al., 2005; Richards and Jia, 2006).

2.2 Decision Tree (DT)

DT is a non-parametric classifier involving a recursive partitioning of the feature space, based on a set of rules learned by an analysis of the training set. It is constructed based on specific rule for each branch involving single or multi-attributes. A new input vector then travels from the root node to a leaf node till it is placed in a specific class (Piramuthu, 2006). Thresholds used for each class decision are chosen using minimum entropy or minimum error measures. It is based on using the minimum number of bits to describe each decision at a node in the tree structure based on the occurrence of each class. DT is terminated with

minimum entropy based on the amount of information gain (i.e. the gain ratio). DT algorithm is stated briefly:

1. If there are N classes denoted $\{C_1, C_2, \ldots C_N\}$, and a training set, T, then
2. If T contains one or more objects which all belong to a single class C_n, then the decision tree is a leaf identifying class C_n.
3. If T contains no objects, the decision tree is a leaf determined from information other than T.
4. If T contains objects that belong to a mixture of classes, then a test is chosen based on a single attribute that has one or more mutually exclusive outcomes $\{O_1, O_2, \ldots, O_N\}$. T is portioned into subsets $T_1, T_2, \ldots, T_n$, where T_n contains all the objects in T that have outcome O_n of the chosen test.

This is done recursively to each subset of training objects to build DT.

2.3 K-Nearest Neighbour (KNN)

Two of the main challenges in RS data interpretation using parametric techniques are high dimensional class data modeling and the associated parameter estimates. Even though the Gaussian normal distribution model has been adopted widely, the need to estimate a large number of covariance terms demand higher number of training samples required for each class of interest. In addition, multimode class data cannot be handled properly with a unimodal Gaussian description. Therefore, non-parametric methods, such as K-Nearest Neighbour (KNN), Neural Network, Random Forest, etc. have the advantage of not needing class density function estimation thereby obviating the training set size problem and the need to resolve multimodality (Han and Kamber, 2003; Venkatesh and Kumar Raja, 2003). The KNN algorithm (Dasarathy, 1990) assumes that pixels close to each other in feature space are likely to belong to the same class. The decision rule is reached directly circumventing the density function estimation. Several decision rules have been developed including a direct majority vote from the nearest k neighbors in the feature space among the training samples, a distance-weighted result and a Bayesian version (Hardin, 1994).

If $\mathbf{x}$ is an unknown pixel vector and suppose there are k_n neighbors labeled as class ω_n out of k nearest neighbours.

$\sum_{n=1}^{N} k_n = k$ (N is the number of classes defined). The basic KNN rule is

$$\mathbf{x} \in \omega_n, \text{ if } m_n(\mathbf{x}) > m_j(\mathbf{x}) \text{ for all } j \neq n \text{ where } m_n(\mathbf{x}) = k_n. \quad (6)$$

If the training data of each class is not in proportion to its respective population, $p(\omega_n)$ in RS data, a Bayesian Nearest-neighbor rule is suggested based on Bayes' theorem

$$m_n(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_n)\, p(\omega_n)}{\sum_{j=1}^{N} p(\mathbf{x}|\omega_n)\, p(\omega_n)} = \frac{k_j p(\omega_n)}{\sum_{j=1}^{N} k_j p(\omega_j)}. \quad (7)$$

The basic rule does not take the distance of each neighbor to the current pixel vector into account and may lead to tied results every now and then. Weighted-distance rule is used to improve upon this as

$$m_n(\mathbf{x}) = \frac{\sum_{j=1}^{k_n} 1/d_{nj}}{\sum_{n=1}^{N} \sum_{j=1}^{k_n} 1/d_{nj}} \quad (8)$$

where d_{nj} is Euclidean distance. With n training samples, k nearest neighbours for every pixel in a large image has to be computed i.e. n spectral distances must be evaluated for each pixel. The above algorithm is summarized as follows:

The variable "unknown" below denotes the number of pixels (count) whose class is unknown and the variable "misclassified" denotes the number of pixels which have been wrongly classified.

set count = 0

set unknown = 0

set misclassified = 0

For all the pixels in the test image

do
{

1. Get the feature vector of the pixel and increment count by 1.
2. From the training set, find the sample feature vector which is nearest neighbour to the feature vector of the pixel.
3. If nearest neighbors count > 1, then check class labels of all the nearest sample feature vectors. If class labels are not the same, then increment unknown by 1 and go to Step 1 to process the next pixel, else go to Step 4.
4. Class label of the image pixel = class label of the nearest sample vector. Go to Step 1 to process the next pixel.

}

2.4 Neural Network (NN)

NN classification has advantages over conventional digital classification algorithms that use spectral distinctiveness of the pixels to decide class labels. The bulk of Multi-layer perceptron (MLP) based classification use multiple layer feed-forward networks that are trained using the back-propagation algorithm based on a recursive learning procedure with a gradient descent search. A detailed introduction can be found in literatures (Atkinson and Tatnall, 1997; Duda et al., 2000; Haykin, 1999; Kavzoglu and Mather, 1999; Kavzoglu and Mather, 2003; Mas, 2003) and case studies (Bischof et al., 1992; Chang and Islam, 2000; Heermann and Khazenie, 1992; Venkatesh and Kumar Raja, 2003).

To train the network for image classification, there are many algorithms. A comparative performance of the training algorithms for image classification by NN is presented in Zhou and Yang, (2010). The MLP in this work is trained using the error back propagation algorithm (Rumelhart et al., 1986). The main aspects here are: (i) the order of training samples should be randomised for each epoch; and (ii) the momentum and learning rate parameters are typically adjusted (generally decreased) with the increase in number of training iterations. Back propagation algorithm for training the MLP is briefly stated below:

i. **Initialize network parameters:** Assign low random real values to weights and biases of the network.
ii. **Present input and desired outputs:** Input a continuous valued vector, $x_0, x_1, \ldots, x_{n-1}$, and specify the desired output $d_0, d_1, \ldots, d_{n-1}$,. If the network is used as a classifier, the desired outputs are set to zero except for that corresponding to the class of the input, which is set to 1.
iii. **Forward computation:** If $[x(n), d(n)]$ denotes a training example in the epoch with $x(n)$ as the input vector applied to the input layer of sensory nodes and $d(n)$ as the desired response vector $d(n)$ presented to the output layer of computation nodes, then the net internal activity $v_j^{(l)}(n)$ for the neuron j in layer l is given by equation (9)

$$v_j^{(l)} = \sum_{i=0}^{p} w_{ji}^{(l)}(n)\, y_i^{(l-1)}(n) \tag{9}$$

where $y_i^{(l-1)}(n)$ is the function signal of neuron i in the previous layer $(l-1)$ at iteration n, and $w_{ji}^{(l)}(n)$ is the synaptic weight of neuron j in the layer l that is fed from neuron i in layer $(l-1)$. Assuming the use of sigmoid function as the nonlinearity, the function (output) signal of neuron j in layer l is given by equation (10)

$$y_i^{(l)} = \frac{1}{1 + \exp[-v_j^{(l)}(n)]} \tag{10}$$

If neuron j is in the first hidden layer (i.e., $l = 1$), set $y_j^{(0)} = x_j(n)$, where $x_j(n)$ is the j^{th} element of the input vector $x(n)$. If neuron j is in the output layer

(i.e., $l = L$), set $y_j^{(L)} = o_j(n)$. Hence, compute the error signal $e_j(n) = d_j(n) - o_j(n)$, where $d_j(n)$ is the j^{th} element of the desired response vector $d(n)$.

iv. **Backward computation:** Compute the δ's (i.e., the local gradients) of the network by proceeding backward, layer by layer:

$$\delta_j^{(l)}(n) = e_j^{(l)}(n)\, o_j(n)[1 - o_j(n)], \text{ for neuron } j \text{ in output layer } L,$$

$$\delta_j^{(l)}(n) = y_j^{(l)}(n)[1 - y_j^{(l)}(n)]$$

$$\sum_k \delta_k^{(l+1)}(n)\, w_{kj}^{(l+1)}(n), \text{ for neuron } j \text{ in the hidden layer } l.$$

Hence adjust the synaptic weights of the network in layer l according to the generalized delta rule (equation 11):

$$w_{ji}^{(l)}(n) = w_{ji}^{(l)} + \alpha[w_{ji}^{(l)}(n-1)] + \eta\, \delta_j^{(l)}(n)\, y_i^{(l-1)}(n) \qquad (11)$$

where η is the learning-rate parameter and α is the momentum constant.

v. **Iteration:** Iterate the forward and backward computations as per steps (iii) and (iv) by presenting new epochs of training examples to the network until stopping criterion is met.

2.5 Random Forest (RF)

RF are ensemble methods using tree-type classifiers $\{h(x, \Theta_k), k = 1, \ldots\ldots,\}$ where the $\{\Theta_k\}$ are i.i.d. random vectors and $\mathbf{x}$ is the input pattern (Breiman, 2001). They are tree predictor combinations where each tree depends on the value of a random vector sampled independently with identical distribution for all trees in the forest. It uses bagging to form an ensemble of classification tree (Breiman, 2001; Gislason et al., 2006). RF is different from other bagging methods in the sense that at each splitting node, a random subset of the predictor variables is used as potential variables to define split. In training, it creates multiple CART (Classification and Regression Tree) trained on a bootstrapped sample of the original training data, and the search is across randomly selected subset of the input variables to determine a split for each node. It utilizes Gini index of node impurity (Breiman, 1998; Breiman et al., 1998) to determine splits in the predictor variables. Each tree casts a unit vote for the most popular class and a majority vote of the trees having highest classification accuracy determines the output.

RF is insensitive to noise and does not overfit. The computational complexity is minimal as trees are not pruned. Therefore, RF can handle high dimensional data, using a large number of trees in the ensemble. Moreover, random selection of variables for a split minimises the correlation between trees in the ensemble, resulting in low error rates that have been compared to those of Adaboost (Freund and Schapire, 1996), at the same time being much lighter in implementation. For more details see (Breiman, 2001; Ham, 2005; Joelsson, 2005; Gislason, 2006;

Prasad, 2006; Walton, 2008; Watts, 2008; Breiman and Cutler, 2010; Na et al., 2010). Some studies have suggested that RF is unexcelled in accuracy among current algorithms (Breiman and Cutler, 2005) and have outperformed CART and similar boosting and bagging-based algorithm (Gislason et al., 2006).

2.6 Contextual Classification using Sequential Maximum a Posteriori Estimation (SMAP)

A training map (obtained from geo-referenced field data) is used to extract spectral signature from RS images. Parameters of a spectral class Gaussian mixture distribution model are determined, which are used for subsequent segmentation (classification) of the data. A class with variety of discrete spectral characteristics is described by a Gaussian mixture. For example, forest, plantation or urban areas are classes that need to be separated in an image. However, each of these classes may contain subclasses each with its own distinctive spectral characteristic; a forest may contain a variety of different tree species each with its own spectral behavior. Mixture classes improve segmentation performance by modeling each information class as a probabilistic mixture with a variety of subclasses.

First, a clustering technique is used to estimate the number of distinct subclasses in each class along with spectral mean and covariance for each subclass. The number of subclasses is estimated using Rissanen's minimum description length (MDL) criteria (Rissanen, 1983). The approximate ML estimates of the mean and covariance of the subclasses are computed using the expectation maximization (EM) algorithm (Dempster, 1977; Redner and Walker, 1984).

The image is segmented into regions at various scales or resolutions using the coarse scale segmentations to guide the finer scale segmentations based on the fact that neighborhood pixels are likely to have similar class (Bouman and Shapiro, 1992; Bouman and Shapiro 1994). The algorithm not only reduces misclassification but also produces segmentation with larger connected regions for a class. However, if nearby pixels often change class, the algorithm adaptively reduces the amount of smoothing, ensuring that excessively large regions are not formed (http://grass.osgeo.org/grass65/manuals/i.smap.html#notes).

2.7 Support Vector Machine (SVM)

SVM are supervised learning algorithms based on statistical learning theory and heuristics (Kavzoglu, and Colkesen, 2009; Yang, 2011). SVM was originally proposed by Vapnik and Chervonenkis (1971) and improvised further by Vapnik (1999). SVM plots input vectors to a higher dimensional space where two parallel hyper planes are constructed on each side of the separating hyper plane that separates the data with lower generalization error. The success of the SVM model depends only on a subset of the training data since training points that lie beyond the margin are unaccounted (Waske and Benediktsson, 2007; Kavzoglu and Colkesen, 2009).

The easiest way to train SVM is by using linearly separable classes (Otukei and Blaschke, 2010).

In order to classify M-dimensional datasets, M-1 dimensional hyper plane is produced with SVMs. As seen in Figure 1, there are various hyper planes separating two classes of data. However, there is only one hyper plane that provides maximum margin between the two classes (Figure 2), which is the optimum hyper plane; the points limiting the width of the margin are the support vectors. With N samples represented by $\{x_n, y_n\}$ (n=1, ..., N), where $x \in R^N$, is an M-dimensional space, and y ε {–1, +1} is class label, then the optimum hyper plane maximizes the margin between the classes.

Figure 2 shows the hyper plane that are defined as $(w \cdot x_n + b = 0)$ (Kavzoglu and Colkesen, 2009), where x is a point lying on the hyper plane, parameter w determines the orientation of the hyper plane in space, b is the bias i.e. the distance of hyper plane from the origin. For the linearly separable case, a separating hyper plane can be defined for two classes as:

$$w \cdot x_n + b \geq +1 \text{ for all } y = +1 \tag{12}$$

$$w \cdot x_n + b \leq -1 \text{ for all } y = -1 \tag{13}$$

where, these inequalities can be combined into a single inequality:

$$y_i(w \cdot x_n + b) - 1 \geq 0 \tag{14}$$

The training data points on these two hyper planes, which are parallel to the optimum hyper plane and defined by the functions $w \cdot x_n + b = \pm 1$, are the support vectors (Mathur and Foody, 2008a). If a hyper plane exists satisfying (14), then the classes are linearly separable. Therefore, the margin between these planes is equal to $2/||w||$ (Mathur and Foody, 2004). As the distance to the closest point is $2/||w||$, the optimum separating hyper plane can be found by minimizing $||w||^2$ under the constraint (14).

The optimum hyper plane is determined by solving

$$\min (0.5^*||w||^2) \tag{15}$$

subject to constraints,

$$y_n (w \cdot x_n + b) \geq -1 \quad \text{and} \quad y_n \varepsilon \{+1, -1\} \tag{16}$$

SVM can function both as a binary classifier and as a multi-class classifier. Figure 3 shows non-linearly separable data as in the classification of remotely sensed images using pixel samples. In such cases, SVM can be extended to allow for non-linear decision surfaces (Cortes and Vapnik, 1995; Pal and Mather, 2003).

In such cases, the optimisation problem is replaced by introducing slack variable ξ.

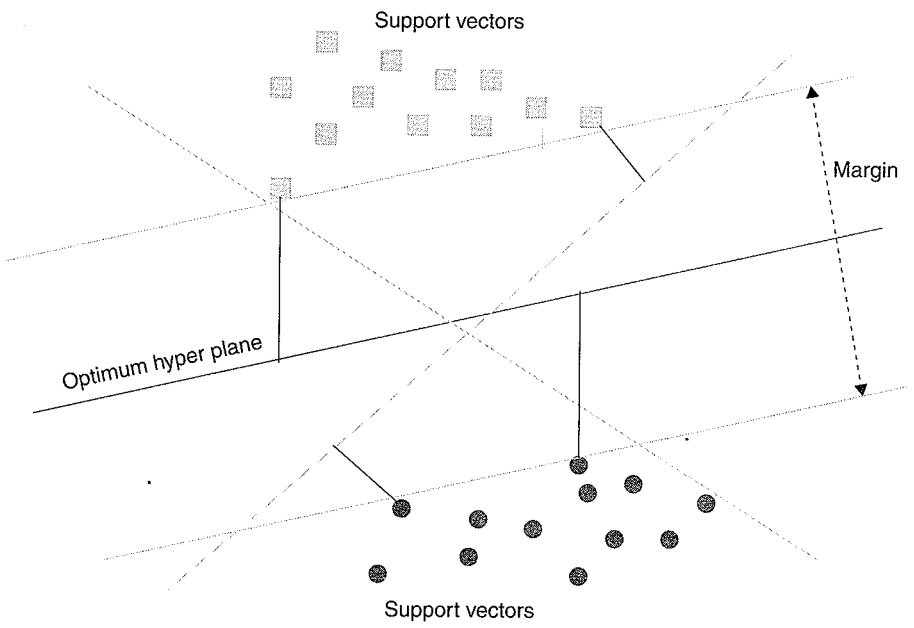

Fig. 1 Optimum hyper plane for linearly separable data with support vectors

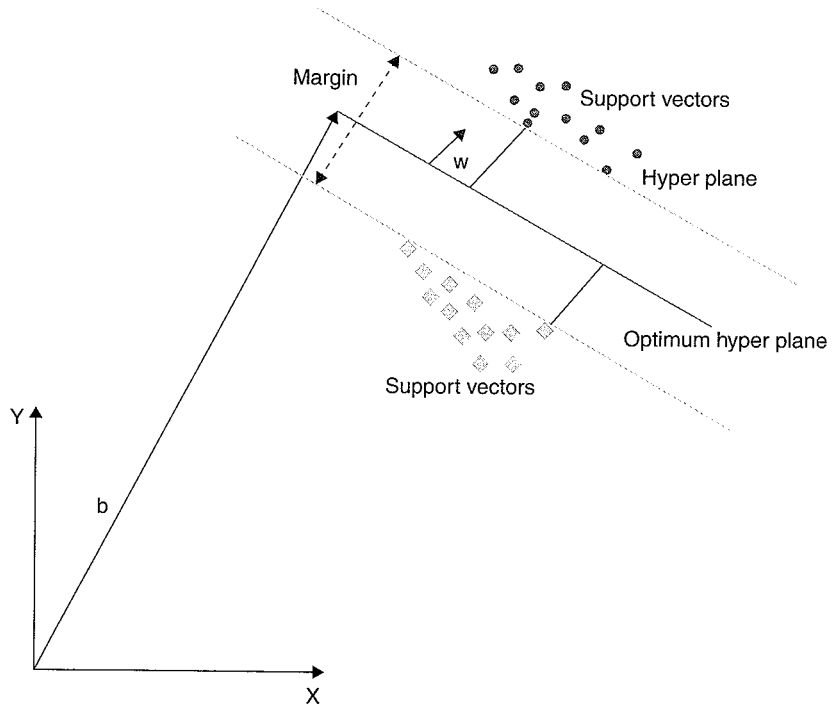

Fig. 2 Optimum hyper plane for the binary linearly separable data

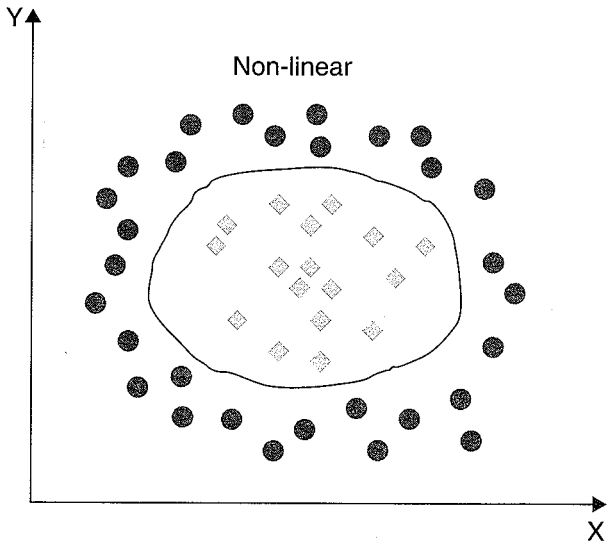

Fig. 3 Separation of non-linear data sets

$$\min\left[\frac{||w||^2}{2} + C\sum_{n=1}^{N}\xi_n\right] \tag{17}$$

subject to constraints

$$y_n(w \cdot x_n + b) \geq 1 - \xi_n, \xi_n, \geq 0, n = 1, ..., N \tag{18}$$

where C is a regularisation constant providing equilibrium between margin maximization and error minimization, and ξ_n indicates distance of the wrongly classified points from the optimal hyper plane (Oommen et al., 2008). Larger the C value, higher is the penalty associated to misclassified pixels (Melgani and Bruzzone, 2004).

When the data cannot be represented linearly, non-linear mapping functions (Φ) are used to map them to a higher dimensional space (H). An input data point x is represented as $\Phi(x)$ in the high-dimensional space. The expensive computation of $(\Phi(x) \cdot \Phi(x_n))$ is reduced by using a kernel function (Mathur and Foody, 2008a). Thus, the classification decision function is

$$f(x) = \text{sign}\left[\sum_{n}^{N} a_n y_n K(x, x_n) + b\right] \tag{19}$$

where for each of N training cases there is a vector ($\mathbf{x}_n$) that is the spectral response with class membership (y_n). a_n ($n = 1, ..., N$) are Lagrange multipliers and $K(x, x_n)$

is the kernel function. The magnitude of a_n is decided by C (Mathur and Foody, 2008b). The kernel function spreads the data points so that a linear hyper plane can be fitted (Dixon and Candade, 2008). Kernel functions commonly used in SVMs are broadly linear, polynomial, radial basis function and sigmoid. The choice of kernel functions and their parameters greatly influence the performance of SV models as we shall observe in this work, where SVM classifier was executed with two types of kernels: polynomial, and radial basis function (RBF):

$$\text{Polynomial: } K(x_i\, x_j) = (\gamma x_i^T x_j + r)^d,\ \gamma > 0 \tag{20}$$

$$\text{RBF: } K(x_i x_j) = \exp(-\gamma || x_i - x_j ||^2),\ \gamma > 0 \tag{21}$$

where, γ is the gamma term for all kernel types except linear, d is the polynomial degree, and r is the bias term. Thecorrect selection of γ, d, and r significantly increase the accuracyof SVM (Chang and Lin, 2001; Wu et al., 2004; Hsu et al., 2007).

3. DATA

The idea of implementing the above classifiers (MLC, DT, KNN, NN, RF, SMAP and SVM) was to investigate their performance on high, medium and low spatial resolution sensor data with unknown distribution. So, IKONOS MS (4 m), IRS LISS (Linear Imaging Self Scanner)-III MS (23.5 m), Landsat ETM+ (Enhanced Thematic Mapper Plus) of 30 m spatial resolution and MODIS (Moderate Resolution Imaging Spectroradiometer) with 7 bands (2 bands at 250 m and 5 bands at 500 m, resampled to 250 m) were chosen in lieu of high, medium and low spatial resolution data (as given in Table 1) because these are the major sensors commonly used for numerous applications in LULC mapping and monitoring. Training and testing data were collected using pre-calibrated hand held GPS (Global Positioning System) for each study area during several field visits and Google Earth data were used to obtain pre-classification information and for post-classification validation.

4. ALGORITHM IMPLEMENTATION

4.1 IKONOS MS

Figure 4 is the high resolution Google Earth image and Figure 5 is the false color composite (FCC) of IKONOS 4 MS bands (700 rows × 700 columns) corresponding to Bangalore City, India. The part of the city shown in scene is highly urbanized with the central business district. It consists of highly contrasting and heterogeneous features such as race course (as oval shape in the first quadrant of the image), bus stand with semi-circular platforms, railway station with railway

Table 1 Remote sensing datasets used for algorithms implementation

Sl. No.	Satellite	Sensor	Date of acquisition	Size	Spectral resolutions	Spatial resolution
1	IKONOS	MS	November 24, 2004)	700 × 700	4 – Blue (B), Green (G), Red (R) and NIR	4 m
2	IRS	LISS-III MS	December 25, 2002	1000 × 1000	R, G and NIR	23.5 m
3	Landsat	ETM + MS	March 14, 2000	2000 × 2000	Band 1 to 5 and band 7 (B, G, R, NIR, MIR-2)	30 m
4	AQUA/ TERRA	MODIS composite 7 bands reflectance	19–26 December, 2002	532 × 546	7 bands (B, G, R, NIR, MIR-2, SWIR)	250 m – (Band 1 and 2); 500 m – (Band 3 to 7) resampled to 250 m

lines in the second quadrant, a park below the race course, dense built up with concrete roofs, and some buildings with asbestos roofs, blue plastic roofs (one in the vicinity of the race course and two near the railway lines), tarred roads with flyovers, vegetation and few open areas (such as a playground, walk ways and vacant land).

For DT, set of rules were extracted using See5 (http://www.rulequest.com) with 25% global pruning. These rules were then used to classify IKONOS MS data. For KNN, number of nearest neighbor was kept 1 in feature space. In case of conflict, random allocation to LU class was done. In NN based classification, a logistic function was used along with 1 hidden layer. Output activation threshold was set to 0.001, training momentum was set to 0.1, training RMS exit criteria were set to 0.1, training threshold contribution was 0.1, and the training rate was maintained at 0.2 to achieve the convergence at 465 iterations (Figure 6). KNN and NN algorithms were coded in C programming language in Linux. RF was implemented using a random forest package (Liaw and Weiner, 2002), available in R interface (http://www.r-project.org). SMAP was implemented through free and open source GRASS GIS (http://wgbis.ces.iisc.ernet.in/grass). SVM was implemented using both polynomial and RBF using libsvm package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). A second degree polynomial kernel was used with 1 as bias in kernel function, gamma as 0.25 (usually taken as 1 divided by the number

of input bands), and penalty as 1. For RBF, gamma was 0.25 and penalty parameter was set to 1.

Classified images of IKONOS data are shown in Figure 7. The class statistics are as listed in Table 2 and Table 3 gives the accuracy assessment with highest 4 overall accuracies highlighted in bold.

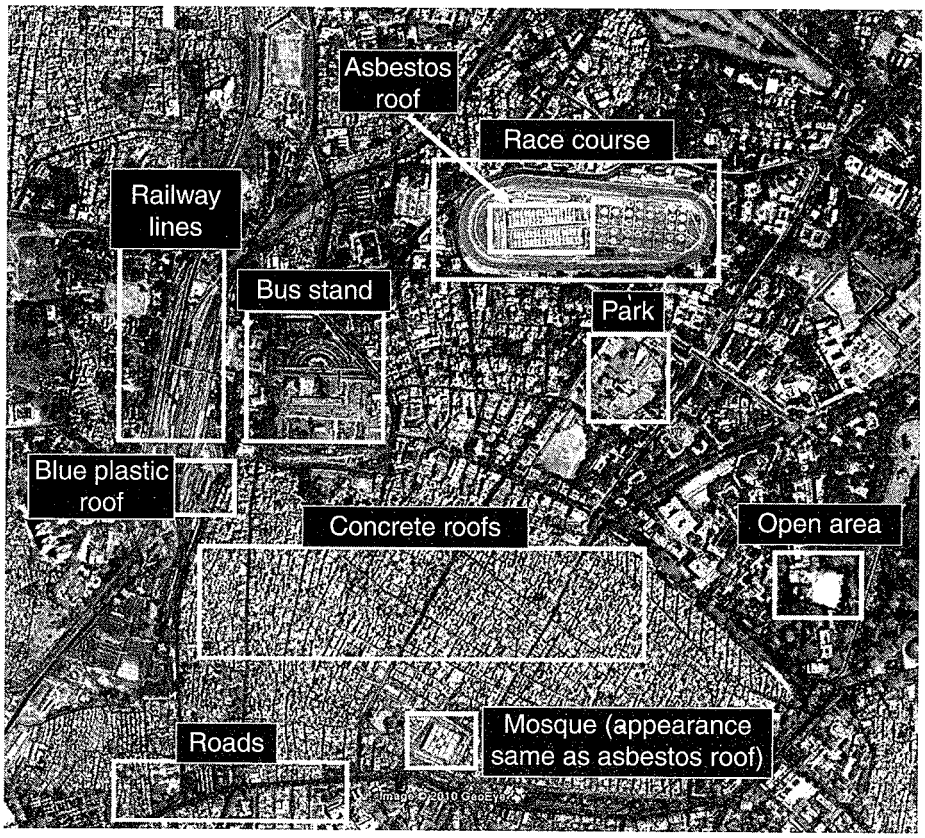

Fig. 4 Google Earth image corresponding to the IKONOS image

4.2 IRS LISS-III MS

LISS-III MS of 23.5 m × 23.5 m spatial resolution (as shown in FCC in the first image in Figure 8 – LISS-III FCC) represents a portion of Kolar district, north of new Bangalore International Airport, India. Training and testing data were collected separately for agriculture, built up, evergreen/semi-evergreen forest, plantations/orchards, wasteland/barren/rock and water bodies. Classified image using the seven algorithms are shown in Figure 8. LU statistics are given in Table 4 and highest 4 overall accuracies are highlighted in bold in Table 5.

Table 2 LU estimates from IKONOS ms using advanced classifiers

Classes →	Concrete roof		Asbestos roof		Blue plastic roof		Vegetation (parks, garden)		Open area (play ground)	
Algorithms↓	ha	%	ha	%	ha	%	ha	%	ha	%
MLC	346.67	44.34	47.99	7.41	5.83	0.75	329.72	42.18	41.60	5.32
DT	344.36	44.05	16.87	2.16	1.86	0.24	259.36	33.18	159.31	20.38
KNN	342.64	43.83	11.77	1.51	11.91	1.52	268.71	34.37	146.73	18.77
NN	381.77	48.83	7.69	0.98	–	–	280.17	35.84	112.12	14.34
RF	351.74	44.99	9.53	1.22	1.88	0.24	260	33.26	158.61	20.29
SMAP	386.36	49.42	4.59	0.59	0.52	0.07	231.71	29.64	158.58	20.28
SVM (Polynom)	361.84	46.28	4.37	0.56	27.34	3.50	246.02	31.47	142.20	18.19
SVM (RBF)	352.83	45.13	4.29	0.55	21.21	2.71	247.38	31.64	156.04	19.96
Total			**781.76 (ha)**						**100 %**	

Fig. 5 FCC of the IKONOS image

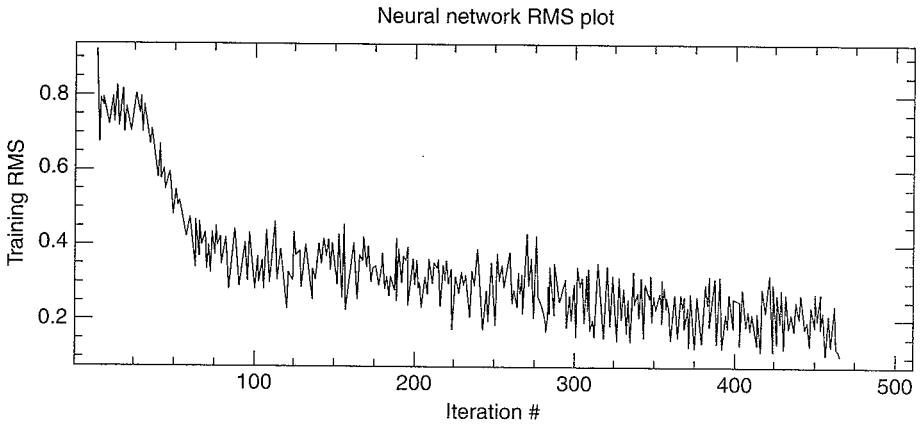

Fig. 6 Plot of training RMS versus iterations of NN

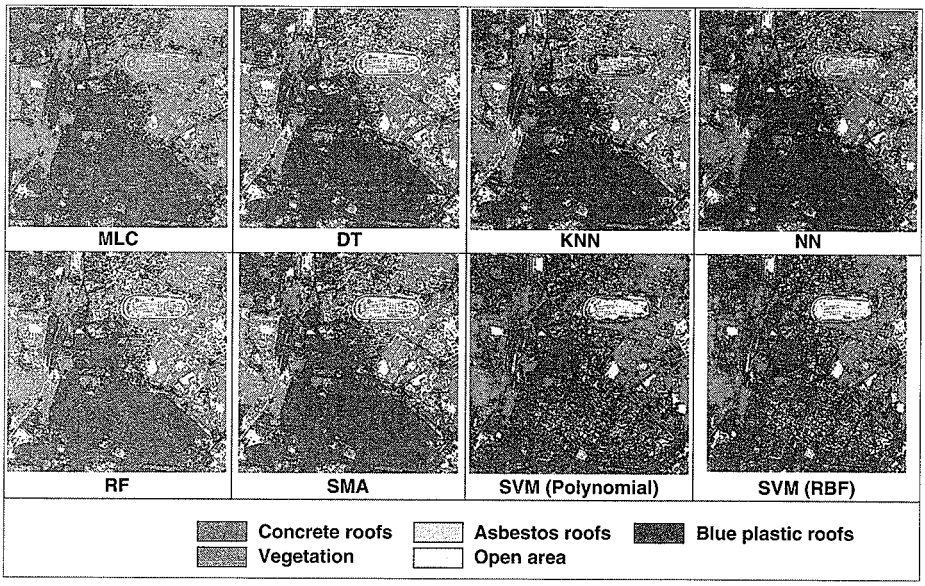

Fig. 7 Classification of IKONOS MS data through advanced classification techniques

Table 3 Accuracy assessment for IKONOS classified data

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
MLC	Concrete	69.99	84.01	**80.46**	**0.6900**
	Asbestos	84.77	87.77		
	Blue Plastic	84.33	81.17		
	Vegetation	94.21	87.55		
	Open area	51.49	69.49		
DT	Concrete	84.74	81.00	**82.22**	**0.8025**
	Asbestos	83.75	79.00		
	Blue Plastic	84.74	80.00		
	Vegetation	82.00	87.00		
	Open area	76.92	83.00		
KNN	Concrete	95.35	95.00	79.85	0.7150
	Asbestos	81.23	75.00		
	Blue Plastic	80.00	79.00		

(*continued*)

Table 3 *Continued*

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
	Vegetation	50.00	75.00		
	Open area	91.00	76.92		
NN	**Concrete**	52.63	68.50	69.84	0.6125
	Asbestos	90.91	50.00		
	Blue Plastic	73.33	71.00		
	Vegetation	71.43	63.00		
	Open area	82.61	75.00		
RF	**Concrete**	92.50	89.99	**85.25**	**0.8250**
	Asbestos	89.15	81.00		
	Blue Plastic	85.00	87.00		
	Vegetation	85.00	83.00		
	Open area	76.92	83.00		
SMAP	**Concrete**	87.21	88.76	**86.92**	**0.8475**
	Asbestos	89.29	83.00		
	Blue Plastic	91.11	88.00		
	Vegetation	91.00	89.00		
	Open area	76.92	85.00		
SVM (Polynomial)	**Concrete**	89.67	55.00	79.97	0.7125
	Asbestos	81.33	60.00		
	Blue Plastic	91.74	85.00		
	Vegetation	95.33	91.00		
	Open area	70.64	80.00		
SVM (RBF)	**Concrete**	89.67	55.00	79.97	0.7125
	Asbestos	81.33	60.00		
	Blue Plastic	91.74	85.00		
	Vegetation	95.33	91.00		
	Open area	70.64	80.00		

4.3 Landsat ETM+ MS

Landsat ETM+ MS data of Uttara Kannada district, Karnataka state, India and part of Western Ghats, India (first image in Figure 9 is Google Earth image) were used for classification. Band 1 to 5 and band 7 of 2000 × 2000 size were classified using training data collected from field and validated using separate test data.

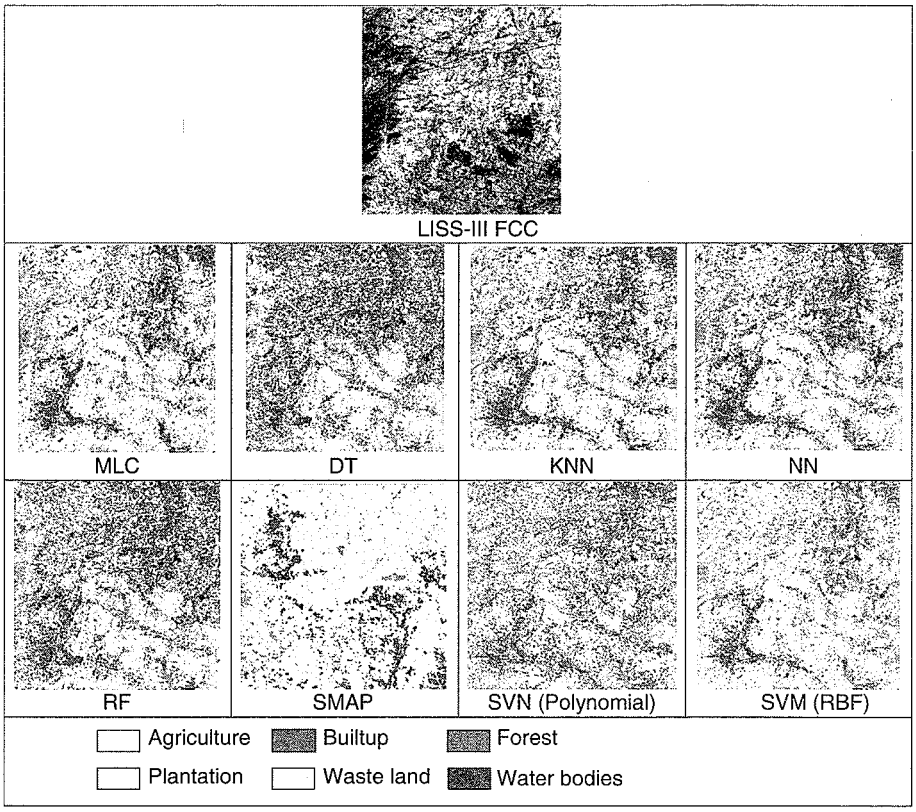

Fig. 8 Classification of IRS LISS-III MS data through advanced classification techniques

Table 4 Accuracy assessment for IRS LISS-III classified data

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
MLC	**Agriculture**	91.21	93.87	**86.59**	**0.8323**
	Builtup	88.33	89.33		
	Forest	94.80	86.24		
	Plantation	88.86	85.46		
	Wasteland	85.20	82.81		
	Water bodies	68.52	85.10		
DT	**Agriculture**	94.72	57.47	75.89	0.7148
	Builtup	63.04	83.60		

(*continued*)

Table 4 *Continued*

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
	Forest	93.87	70.38		
	Plantation	95.75	76.15		
	Wasteland	56.91	78.37		
	Water bodies	76.38	67.00		
KNN	Agriculture	89.36	89.44	**89.02**	**0.8604**
	Builtup	87.54	95.66		
	Forest	94.90	87.99		
	Plantation	89.00	88.00		
	Wasteland	85.01	82.32		
	Water bodies	85.18	94.00		
NN	Agriculture	92.56	81.45	**83.66**	**0.8112**
	Builtup	81.30	91.63		
	Forest	87.90	88.34		
	Plantation	83.00	81.00		
	Wasteland	74.42	81.56		
	Water bodies	78.83	82.00		
RF	Agriculture	93.96	74.73	82.29	0.7638
	Builtup	67.51	83.65		
	Forest	91.57	85.70		
	Plantation	80.00	81.00		
	Wasteland	73.47	84.87		
	Water bodies	85.09	86.00		
SMAP	Agriculture	91.17	62.46	76.17	0.7365
	Builtup	79.49	64.98		
	Forest	73.23	87.84		
	Plantation	75.00	81.00		
	Wasteland	51.15	72.56		
	Water bodies	88.18	87.00		
SVM (Polynomial)	Agriculture	75.28	48.04	68.83	0.6647
	Builtup	33.63	50.17		
	Forest	90.55	84.49		
	Plantation	64.00	68.00		
	Wasteland	53.71	92.73		
	Water bodies	69.35	96.00		

(*continued*)

Table 4 *Continued*

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
SVM (RBF)	Agriculture	90.49	85.83	88.18	0.8534
	Builtup	85.54	89.33		
	Forest	90.90	86.69		
	Plantation	93.14	87.15		
	Wasteland	82.46	85.51		
	Water bodies	90.17	91.00		

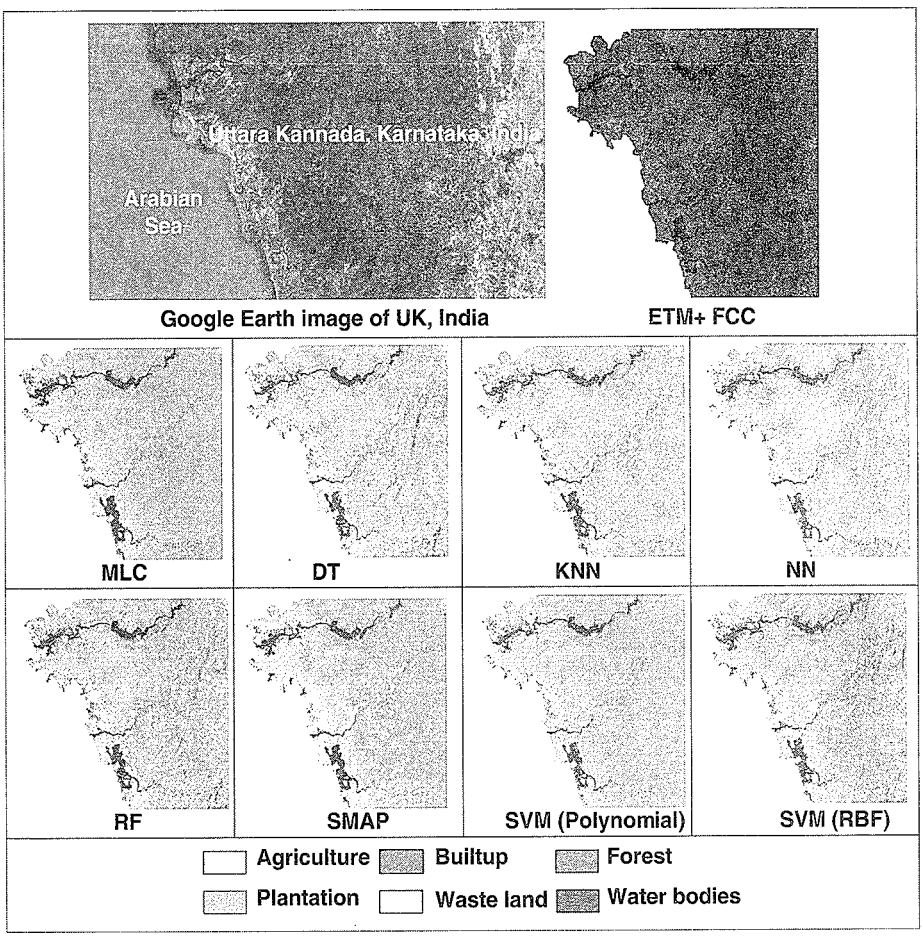

Fig. 9 Classification of ETM+ Plus data through advanced classification techniques

Table 5 LU estimates from IRS LISS-III MS using advanced classifiers

Classes →	Agriculture		Built up		Forest		Plantation		Wasteland		Water bodies	
Algorithms ↓	ha	%	ha	%	ha	%	ha	%	ha	%	ha	%
MLC	15388	32.32	7524	15.80	4091	8.59	4951	10.40	15184	31.89	473.13	0.99
DT	10510	21.94	13923	29.07	1307	2.73	12210	25.49	9467	19.76	483.57	1.01
KNN	20447	42.69	10109	21.10	2320	4.84	5993	12.51	8797	18.37	235.95	0.49
NN	18455	38.53	9856	20.58	3619	7.56	3298	6.88	12082	25.22	591.34	1.23
RF	16016	33.44	12916	26.96	2492	5.20	5973	12.47	9990	20.85	514.52	1.07
SMAP	19431	40.56	4279	6.93	1029	2.15	4968	10.37	18057	37.70	137.72	0.29
SVM (Poly)	14360	29.87	11830	24.70	1227	2.56	4862	10.15	15610	32.59	66.72	0.14
SVM (RBF)	20071	41.9	6571	13.72	2709	5.65	7860	16.41	10488	21.89	203.3	0.42
Total	47901.00 (ha)								100%			

Table 6 LU estimates from ETM+ using advanced classifiers

Classes →	Agriculture		Builtup		Forest		Plantation		Wasteland		Water bodies	
Algorithms↓	ha	%	ha	%	ha	%	ha	%	ha	%	ha	%
MLC	18824	5.80	6386	1.97	226695	69.88	54043	16.66	5322	1.64	13157	4.06
DT	51617	15.91	6398	1.97	158049	48.72	83446	25.72	9411	2.90	15502	4.78
KNN	48566	14.97	8663	2.67	195457	60.25	50793	15.66	7582	2.34	13361	4.12
NN	39075	12.04	8068	2.49	151979	46.85	97686	30.11	10609	3.27	17005	5.24
RF	41629	12.83	5732	1.77	194666	60.00	56483	17.41	10826	3.34	15121	4.66
SMAP	35992	11.09	4249	1.34	201454	62.01	64034	19.74	4953	1.53	13739	4.24
SVM (Poly)	19292	5.95	6385	1.97	226815	69.92	54024	16.65	5319	1.64	12548	3.87
SVM (RBF)	37680	11.62	6421	1.98	193195	59.56	54437	16.78	5319	1.64	27331	8.43
Total			324421.68 (ha)						100%			

Table 7 Accuracy assessment for ETM+ classified data

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
MLC	Agriculture	82.50	86.66	**85.18**	**0.8190**
	Builtup	84.00	85.00		
	Forest	80.85	90.91		
	Plantation	81.66	83.00		
	Wasteland	83.00	89.67		
	Water bodies	90.91	84.00		
DT	Agriculture	83.33	86.67	84.54	0.7946
	Builtup	95.00	85.00		
	Forest	82.63	80.61		
	Plantation	85.75	80.00		
	Wasteland	85.00	84.00		
	Water bodies	83.33	83.21		
KNN	Agriculture	84.41	87.00	**86.98**	**0.8314**
	Builtup	97.00	87.00		
	Forest	76.43	89.79		
	Plantation	87.45	86.67		
	Wasteland	87.00	89.00		
	Water bodies	87.00	85.00		
NN	Agriculture	83.33	70.00	74.98	0.7142
	Builtup	85.00	82.00		
	Forest	62.63	60.61		
	Plantation	87.85	76.66		
	Wasteland	77.00	71.00		
	Water bodies	66.67	77.00		
RF	Agriculture	87.44	86.66	83.37	0.7505
	Builtup	87.00	82.00		
	Forest	74.26	81.82		
	Plantation	82.57	84.73		
	Wasteland	82.00	79.00		
	Water bodies	90.91	82.00		
SMAP	Agriculture	85.48	86.66	**89.03**	**0.8596**
	Builtup	98.00	99.00		

(*continued*)

Table 7 *Continued*

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
	Forest	80.65	87.76		
	Plantation	88.94	89.57		
	Wasteland	89.00	87.00		
	Water bodies	87.33	89.00		
SVM (Polynomial)	**Agriculture**	87.27	80.00	**85.35**	**0.8324**
	Builtup	85.00	95.00		
	Forest	88.70	81.82		
	Plantation	81.66	85.47		
	Wasteland	85.00	87.00		
	Water bodies	85.55	81.67		
SVM (RBF)	**Agriculture**	76.25	83.33	**83.77**	**0.7977**
	Builtup	80.00	93.00		
	Forest	80.85	80.91		
	Plantation	81.66	83.33		
	Wasteland	89.00	85.00		
	Water bodies	90.91	81.00		

The classified images using the seven techniques are shown in Figure 9. Tables 6 and 7 give the LU statistics and accuracy assessment with highest 4 accuracies highlighted in bold.

4.4 MODIS

MODIS 8-day composite data (7 bands of 250 m × 250 m) were co-registered with known ground control points – GCPs (RMSE – 0.11) representing Kolar district in Karnataka having an area of 8238 sq. km. Training data (≥ 250 m × 250 m) were collected representing approximately 10% of the geographical area covering the entire spectral gradient of the classes. Separate test data were collected for validation. Preliminary survey revealed that there are six major LU classes in the study area – agriculture, builtup (urban/rural), evergreen/semi-evergreen forest, plantations/orchards, wasteland/barren rock/stony waste/sheet rock and water bodies/lakes/ponds/tanks/wetlands) that could constitute homogeneous MODIS pixels. Barren/rock/stone that have very limited ground area proportions and are unevenly scattered among the major six classes, were grouped under the wasteland category since they would form mixed pixels.

The classified images are shown in Figure 10 along with LU statistics (Table 8) and accuracy assessment (in Table 9).

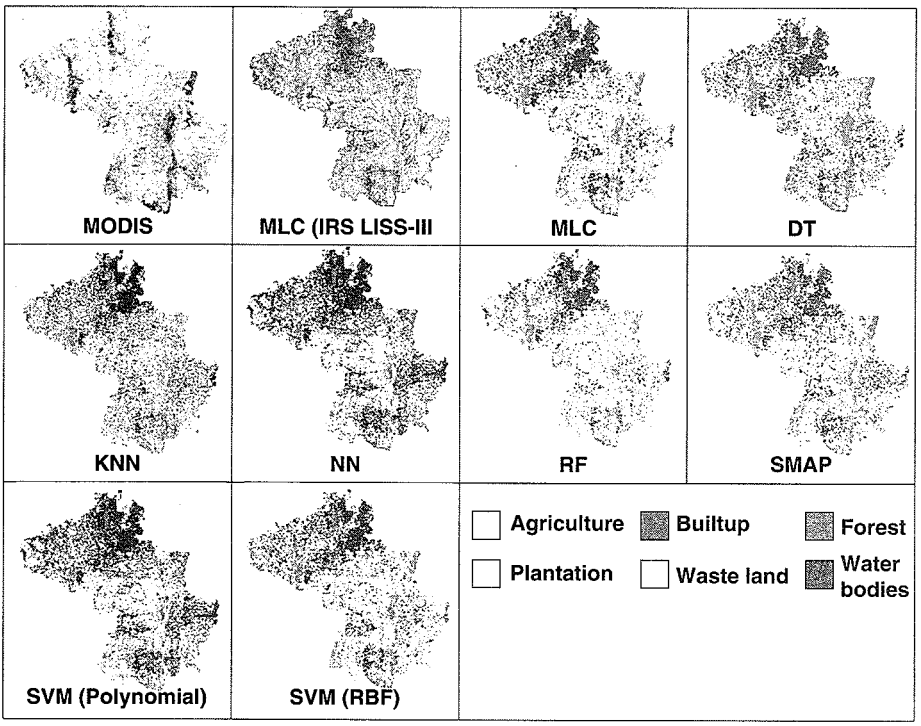

Fig. 10 Classification of MODIS data through advanced classification techniques

5. DISCUSSION

The performance of each classifier is discussed in this section, with reference to IKONOS MS, IRS LISS-III MS, Landsat ETM+ MS and MODIS data.

From visual interpretation and comparison of the IKONOS classified image with the Google Earth and ground truth data, it was noticed that among the seven techniques used for classifying IKONOS images, SMAP performed best with 86.92% overall accuracy followed by RF with 85.25% accuracy (Table 10). SMAP takes into account the intra class spectral variations and exploits spatial information among neighboring pixels to improve classification results (Magnussen et al., 2004). NN was difficult to train before it reached convergence as evident from the training RMS plot in Figure 6. It has wrongly classified asbestos roof to concrete built-up and was unable to detect the blue plastic roof (Table 2). SMAP has aggregated some pixels of similar class types, creating blocky appearance as

Table 8 LU estimates from MODIS using advanced classifiers

Classes →	Agriculture		Builtup		Forest		Plantation		Wasteland		Water bodies	
Algorithms↓	ha	%	ha	%	Ha	%	ha	%	ha	%	ha	%
MLC	170247	20.68	130650	15.87	85352	10.37	107970	13.12	320767	38.57	8156	0.99
DT	126752	15.92	133521	16.77	123061	15.46	93911	11.80	314126	39.46	4790	0.60
KNN	1702834	20.69	142703	17.34	99731	12.12	99441	12.08	302091	36.70	8892	1.08
NN	179720	21.83	217984	26.48	629012	7.64	160886	19.55	199462	24.23	2186	0.27
RF	238517	28.99	97328	11.83	66366	8.07	54214	6.59	363934	44.23	2392	0.29
SMAP	224464	27.27	117323	14.25	79338	9.64	83048	10.09	320813	37.76	8156	0.99
SVM (Poly)	180567	21.94	215810	26.22	63072	7.66	161386	19.61	199785	24.28	2382	0.29
SVM (RBF)	171020	20.80	127450	15.50	85565	10.41	108378	13.18	321455	39.11	8162	0.99
Total			823141.42 (ha)						100%			

Table 9 Accuracy assessment for MODIS classified data

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
MLC	Agriculture	93.70	70.26	**70.43**	**0.6721**
	Builtup	44.00	46.32		
	Forest	80.00	66.22		
	Plantation	65.90	77.88		
	Wasteland	68.90	83.33		
	Water bodies	70.75	75.38		
DT	Agriculture	60.24	53.07	59.51	0.4860
	Builtup	58.06	63.00		
	Forest	38.02	58.35		
	Plantation	62.00	54.00		
	Wasteland	65.19	63.28		
	Water bodies	71.00	67.85		
KNN	Agriculture	53.82	95.56	62.74	0.5801
	Builtup	48.90	44.33		
	Forest	65.00	64.00		
	Plantation	58.18	63.85		
	Wasteland	71.31	45.86		
	Water bodies	70.87	71.29		
NN	Agriculture	54.44	76.08	64.14	0.5633
	Builtup	48.73	44.33		
	Forest	64.00	63.87		
	Plantation	83.04	75.39		
	Wasteland	65.00	58.62		
	Water bodies	68.76	67.44		
RF	Agriculture	62.98	88.17	**76.21**	**0.7357**
	Builtup	59.44	54.33		
	Forest	75.00	79.00		
	Plantation	92.93	91.03		
	Wasteland	83.17	72.41		
	Water bodies	79.00	77.00		
SMAP	Agriculture	67.26	71.69	**69.44**	**0.6558**
	Builtup	47.83	54.00		
	Forest	82.54	75.43		
	Plantation	61.70	65.72		

(*continued*)

Table 9 *Continued*

Algorithm	Class	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
	Wasteland	78.87	71.17		
	Water bodies	81.47	75.56		
SVM (Polynomial)	**Agriculture**	54.44	76.08	**64.65**	**0.5633**
	Builtup	48.73	44.33		
	Forest	64.00	61.00		
	Plantation	83.04	85.39		
	Wasteland	67.00	58.62		
	Water bodies	68.41	64.70		
SVM (RBF)	**Agriculture**	70.26	78.70	**70.75**	**0.6921**
	Builtup	56.32	57.34		
	Forest	66.22	73.34		
	Plantation	77.88	65.90		
	Wasteland	83.33	68.10		
	Water bodies	72.43	79.22		

evident from the disappearance of the linear tarred road (flyover), merging with the adjacent concrete built-up. Inside the race course area, asbestos roof is mixed with open area in MLC classified image, blue plastic roof has been overestimated in KNN and SVM (Poly), open area is overestimated among the concrete buildings especially in race course in SVM (RBF) as seen in Figure 7, bringing down its user's accuracy. SVM (Poly) and SVM (RBF) have similar accuracies.

LISS-III data belongs to a region where forest and plantation are mixed because of degradation of natural forest and afforestation as plantation. Forest has exposed soil between trees and is therefore not dense to have spectrally significant different signature compared to plantation. As a result, the two classes often mix and create confusion in interpreting classification result. The area has dominant wasteland which has similar reflectance as built up or barren land. Wasteland has often been misclassified as built up and vice-versa. KNN has classified LISS-III imagery with maximum accuracy (Table 5). Agriculture is overestimated in SMAP and much wasteland is misclassified as built up in DT and RF (Figure 8). Plantation class has increased in DT due to mislabeled forest pixels. SMAP and KNN have problems in classifying water bodies and have underestimated the same. The second best classifier followed by KNN (89.02%) in first position is SVM (RBF) with 88% accuracy. A NN with 4 hidden layers, 0.1 learning rate and 0.1 momentum with 5000 epochs was used for training the network in 14 seconds which was fourth best classifier with 83.66% overall accuracy after MLC (86.59%) in third position.

Landsat data having a spatial resolution of 30 m were classified most accurately using SMAP algorithm (89.03% overall accuracy, Table 7). The area covered by this image has forested landscape, dominated by evergreen and semi-evergreen flora. KNN, SVM (Poly) and MLC gave higher overall accuracies of more than 85%. Plantation was overestimated in DT and NN which had 4 hidden layers with 0.2 learning rate and 0.2 momentum with 20000 epochs, took 77 seconds to train. A second degree polynomial function with gamma as 0.167 was used in SVM (Poly), which gave lower accuracies in detecting water bodies in comparison to other seven techniques. DT, NN, RF and SVM (RBF) showed abnormal trends and have classified mountain ridges as narrow water channels (Figure 9). Wasteland are often mixed with fallow land due to seasonal differences in crop practices, and it also reflects similar to sand on the sea shores/sea beaches (as depicted in first picture in Figure 9: Google Earth image of UK, India with Arabian sea on the west portion of the image), which were prominent in DT classification.

Due to coarse spatial resolution of MODIS pixels (250 m), misclassification seems to be obvious because of mixed pixel problem that was unavoidable. With pixilated and patchy appearance of classified images, overall classification accuracy ranged from 59.5% (DT) to 76.21% (RF). When compared with the ground truth and resampled LISS-III MS classified image at 25 m spatial resolution, (Figure 10), RF has performed best in classifying LU categories, followed by SVM (RBF), MLC and SMAP. DT has underestimated plantation, while NN has classified many wasteland pixels as built up in the northern part of study area. Whatever be the performance of each technique, the classified MODIS images are not useful in obtaining accurate LULC information at regional level. Table 10 provides the best four classification techniques for various resolution data depending on their performance and overall accuracies.

Table 10 illustrates that SMAP, KNN, SVM and RF are applicable at regional level LULC mapping using medium to high spatial resolution data. RF has outperformed DT and is also reported by Gislason et al., (2006). RF are increasingly being merged with other techniques such as object oriented approaches for achieving better classification accuracy (Watts and Lawrence, 2008), for

Table 10 Best performing classification algorithms for different sensors

Rank	IKONOS	OA*	LISS-III	OA*	ETM	OA*	MODIS	OA*
1	SMAP	86.92	KNN	89.02	SMAP	89.03	RF	76.21
2	RF	85.25	SVM (RBF)	88.18	KNN	86.98	SVM (RBF)	70.75
3	DT	82.22	MLC	86.59	SVM (Poly)	85.35	MLC	70.43
4	MLC	80.46	NN	83.66	MLC	85.18	SMAP	69.44

*OA – Overall Accuracy (%)

hyperspectral data classification (Ham et al., 2005; Joelsson et al., 2005), for ecological prediction (Prasad et al., 2006), etc.

However, the results obtained here are contrary to Otukei and Blaschke, (2010), who indicated that DT was better than SVM and MLC. SVM has provided promising results in LULC classification (Watanachaturaporn, et al., 2004; He et al., 2005; Kavzoglu and Colkesen, 2009; Watanachaturaporn, et al., 2007, 2008). In spite of many advanced classifiers that do not depend on the underlying data distribution, MLC, which works on the assumption of normality, has performed better than many other techniques.

This analysis showed that overall classification accuracy of the IKONOS images were between 69.84 and 86.92%. The highest overall accuracy obtained from SMAP (86.92%) with IKONOS bands is still lower than highest overall accuracy from IRS LISS-III (89.02%) using KNN, and 89.03% from SMAP with ETM+ image classification. One reason is that urban areas contain highly contrasting features that exhibit similar reflectance's, creating spectral confusion between classes, such as building roofs of different types. Even with very high resolution data such as IKONOS MS, classification is not always easy and leads to poor accuracy in areas with mixture of land uses. ETM+ image classification ranged between 74% and 89% which are comparable to LISS-III image classification with reasonably good accuracies for mapping LU dynamics at a regional scale. These results are parameter specific and may not give similar output when the parameters are altered or adjusted. Classification errors in MODIS have occurred since the signal of the pixels are ambiguous or perhaps as a result of spectral mixing in pixels. Also, as the pixel resolution becomes coarser (in this case 250 m), the chance of high accuracy as the product of random assignment of values also declines. On similar lines, Quattrochi and Goodchild (1997) suggest that a fine-scale classification system is needed for a classification at local level. Thus high spatial resolution data such as IKONOS and SPOT 5 HRG are helpful. At a regional scale, medium spatial resolutions such as Landsat TM/ETM+, Terra ASTER are most frequently used data. At a continental or global scale, coarse spatial resolution data such as AVHRR, MODIS, and SPOT Vegetation are preferable. However, a recent study by Eva et al., (2010) demonstrates the usage of medium spatial resolution satellite imagery (Landsat-5 TM and SPOT-HRV) for monitoring forest areas from continental to territorial levels. Inter-sensor comparison between Resourcesat, LISS-III, LISS-IV and AWiFS with reference to coastal LULC (mudflat, mangroves, vegetated dune, coastal water, etc.) were carried out on the basis of DN values, converting radiance and reflectance values of each sensor (Chauhan and Dwivedi, 2008). The study revealed that LISS-IV can be used in place of LISS-III or it can be merged (LISS-III MS + PAN). Comparison of AWiFS and LISS-IV did not show good correlation as with LISS-III, because of the large difference in spatial resolutions.

Spatial resolution is an important factor that affects classification details and accuracy (Chen et al., 2004; Velpuri et al., 2009) and influences the selection of classification approaches (Atkinson and Aplin, 2004), since the size of ground

objects relative to the spatial resolution of a sensor is directly related to image variance (Woodcock and Strahler, 1987). When the object in the scene becomes increasingly smaller relative to the pixel resolution, they are no longer regarded as individual objects. In such cases, reflectance is treated as a sum of interactions among various classes as weighted by their relative proportions (Strahler et al., 1986). Therefore, medium spatial resolution data such as IRS LISS-III and Landsat are useful at a regional scale but not appropriate at a local level.

For fine spatial resolution data, although mixed pixels are reduced, the spectral variations within classes decrease the classification accuracy (Lu and Weng, 2007). Therefore, selection of a suitable data mining algorithm must consider important factors such as aim of classification, classification accuracy, algorithm performance, computational resources and effective separation of classes. Lu et al., (2004) and South et al., (2004) argue that in many cases, contextual-based classifiers, per-field approaches, and machine-learning approaches provide a better classification result than MLC, although some tradeoffs exist in classification accuracy, time consumption, and computing resources. A combination of multi-sensor data with varying image characteristics, considering the economic conditions, are also important factors that affect the selection of RS data, classification technique and time affecting the quality of classification results.

Uncertainties involved in different stages of classification procedures influence classification accuracy, as well as the area estimation under different LULC classes (Dungan, 2002). Understanding relationships between classification stages, identifying the probable factors influencing the accuracy and improving them are essential for successful image classification. For example, spatial or radiometric resolution constraints, geometric rectification of multi-sensor data and the atmospheric condition during image acquisition time causes uncertainty. On the other hand, the algorithm used for calibrating atmospheric or topographic effects may cause radiometric errors (Lu and Weng, 2007). Dungan (2002) listed five types of uncertainties in RS data processing such as positional, support, parametric, structural (model) and variables. Friedl et al., (2001), emphasises on three sources of errors, namely, image acquisition process, during data processing and the interaction between resolution and the scale of ecological processes. Yu et al., (2008) has discussed the factors affecting spatial variation of classification uncertainty in an image object-based vegetation mapping. Such errors have to be considered while handling coarse resolution data such as MODIS, due to existence of many mixtures among LULC classes. In such cases, geocomputation (Beekhuizen and Clarke, 2010), geo-visualisation and interactive visualisation techniques (Lucieer and Kraak, 2004) have proved to be useful.

Spectral characteristic of RS data needs to be considered during classification using data mining algorithms. As spatial resolution increases, texture or contextual information play vital role and turn out to be decisive factors. Classification approaches may vary with varying RS data. For instance, with high spatial resolution data such as IKONOS, the impact of shadow resulting from topography and trees and wide spectral variation within the LU classes may outweigh the advantages of high spatial resolution during classification (Lu and Weng, 2007),

where contextual based classification such as SMAP may be better as evident from the present study (Table 3). In this case, a combinatorial approach of spectral and textural information can reduce the problem. For medium to coarse resolution data, however, spectral information is more important than spatial where per-pixel classification may not be appropriate. In such cases, image fusion or sub-pixel classification would be more suitable. For a particular study, it is often difficult to identify the best classifier due to lack of a guideline for selection, and the availability of suitable classification algorithm. Comparative study of different classifiers such as the one conducted here, would certainly add to the already existing information on spatio-spectral image classification. Next, we discuss a new Hybrid Bayesian Classifier.

6. HYBRID BAYESIAN CLASSIFIER (HBC)

HBC is based on linear unmixing and Bayesian classifier, to assign a class label to each pixel in a *high spatial-low spectral resolution* (HS-LSR) MS data (Kumar et al., 2011). The prior probabilities of the different classes in the Bayesian classifier to classify each pixel in the HS-LSR data such as IRS LISS-III MS and IKONOS MS are obtained from abundance estimates by unmixing *low spatial-high spectral resolution* (LS-HSR) MS data such as MODIS and Landsat ETM+ respectively. The terms HS-LSR and LS-HSR are relative, depending upon the spatial resolution of the images. For instance, MODIS bands are referred as LS-HSR and IRS LISS-III MS bands are HS-LSR data, while Landsat ETM+ bands are LS-HSR and IKONOS MS bands are HS-LSR data. The reason for selecting MODIS and Landsat ETM+ images as LS-HSR supplement data is because of their economic viability and high temporal frequency that enables their procurement for any part of the globe, throughout the year corresponding to any HS-LSR data. However, the technique in general, can be applied on any other LS-HSR (such as hyperspectral bands) and HS-LSR (such as multispectral bands) data classification. The novelty of this approach lies in the fact that low spatial-high spectral and low spectral-high spatial resolution MS data are combined to improve the classification results which can be thought of as a fusion process, in the sense, that information from two different sources (sensors) are combined to arrive at an improved classified image by systematically exploiting the relevant information from both the sources as shown in Figure 11.

In the following section, linear unmixing using Orthogonal Subspace Projection (OSP) and Bayesian classifier are discussed. Next, HBC is discussed with experimental results and validation.

6.1 Linear Unmixing

With M spectral bands and N classes (C_1, .., C_N), each pixel has an associated M-dimensional pixel vector $y = (y_1, \ldots, y_M)$ whose components are the pixel

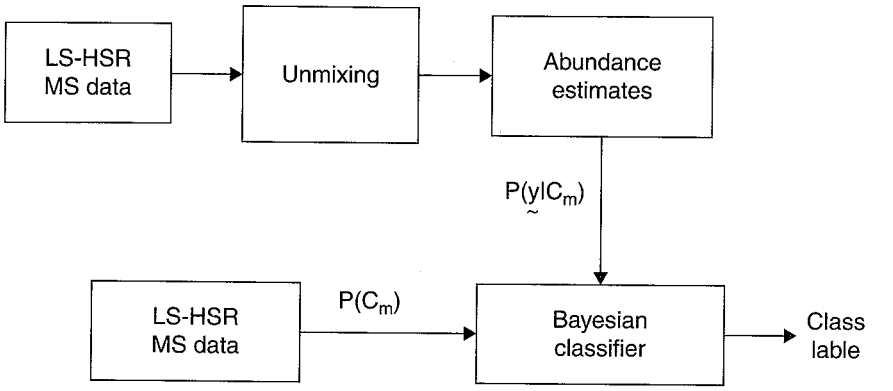

Fig. 11 Hybrid bayesian classifier

intensities corresponding to the M spectral bands. Let $\boldsymbol{E} = [e_1, \ldots, e_N]$, where, for $n = 1, \ldots, N$, e_n is a $M \times 1$ column vector representing the end member spectral signature of the n^{th} target material. For a pixel, let a_n denote the fraction of the n^{th} target material signature, and $a = (a_1, \ldots, a_N)$ denote the N-dimensional abundance column vector. The linear mixture model for y is given by

$$\boldsymbol{y} = \boldsymbol{E}a + \eta \tag{22}$$

where, $\eta = (\eta_1, \ldots, \eta_M)$, are i.i.d. $N(0, \sigma^2)$ (Kumar et al., 2008). Equation (22) represents a standard signal detection model where $\boldsymbol{E}a$ is a desired signal vector to be detected. Since, OSP detects one target at a time, we divide a set of N targets into desired (C_n) and undesired $(C_n, \ldots, C_{n-1}, C_{n+1}, \ldots C_N)$ targets. A logical approach is to eliminate the effects of undesired targets that are considered as "impeders" to C_n before detecting C_n To estimate a_n, the desired target material is e_n,. The term Ea in (22) can be rewritten to separate the desired spectral signature e_n,:

$$\boldsymbol{y} = e_n a_n + \boldsymbol{R}\boldsymbol{r} + \eta \tag{23}$$

where $\boldsymbol{r} = (a_1, \ldots a_{n-1}, a_n+_1, \ldots, a_N)$ and $\boldsymbol{R} = [e_1, \ldots e_{n-1}, e_n+_1, \ldots, e_N]$ Thus the interfering signatures in $\boldsymbol{R}$ can be removed by the operator,

$$\boldsymbol{P} = .(\boldsymbol{I} - \boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{R})^{-1}\boldsymbol{R}^T) \tag{24}$$

which is used to project y into a space orthogonal to the space spanned by the interfering spectral signatures, where I is the $K \times K$ identity matrix (Change, 2005). Operating on y with $\boldsymbol{P}$, and noting that $\boldsymbol{PR} = 0$,

$$\boldsymbol{P}\boldsymbol{y} = Pe_n a_n + \boldsymbol{P}\eta \tag{25}$$

After maximizing the SNR, an optimal estimate of a_n is

$$a_n = \frac{y^T P^T P y}{e_n^T P^T P e_n} \tag{26}$$

Note that $1 \geq a_n \geq 0$ and thus $(a_1, \ldots, a_N)$ can be taken as proportional probabilities of the N classes i.e. $P(C_n) \propto a_n$.

6.2 Bayesian Classifier

Associated with any pixel, there is an observation y With N classes $(C_1, \ldots, C_N)$, Bayesian classifier calculates the posterior probability of each class conditioned on y (Hahn and Kamber):

$$P(C_n \mid y) = \frac{P(y \mid C_n)\, P(C_n)}{P(y)} \tag{27}$$

In (27), since $P(y)$ is constant for all classes, only $P(y|C_n)P(C_n)$ is considered. $P(y \mid C_n)$ is computed assuming class conditional independence, so, $P(y \mid C_n)$ is given by

$$P(y \mid C_n) = \prod_{m=1}^{M} P(y_m \mid C_n) \tag{28}$$

6.3 Hybrid Bayesian Classifier (HBC)

HBC uses the abundance of each class obtained from LS-HSR data by linear unmixing as prior probability while classifying the HS-LSR data using Bayesian classifier of the same geographical area and time frame. That is, given the observation vector y for a pixel, it is classified to fall in class l if l = Arg Max $P(y \mid C_n)$ $P(C_n)$, where $P(y \mid C_n)$ is as in (28) calculated using the HS-LSR data and $P(C_n) \propto a_n$ calculated using the LS-HSR data. The assumptions are: (i) if there are r HS-LSR pixels contained in one LS-HSR pixel, i.e. resolution ratio is $(r : 1)$, the prior probabilities for all the r HS-LSR pixels are equal corresponding to the same LS-HSR pixel, and (ii) the two data types have a common origin or upper left corner, i.e. the edges of the $r \times r$ HS-LSR pixels overlaps exactly with the corresponding LS-HSR pixel. The limitations are: (i) (K-1) should be $\geq$ M in LS-HSR data and (ii) M in HS-LSR should be $\leq$ M in LS-HSR data.

6.4 Experimental Results

Two separate experiments were carried out. In the first experiment, LISS-III MS data (3 bands of 23.5m × 23.5m spatial resolution resampled to 25m, acquired on December 25, 2002) of 5320 × 5460 size and MODIS 8-day composite data

(7 bands of 250m × 250m, acquired from 19–26 December, 2002) of 532 × 546 dimension were co-registered with known ground control points (RMSE – 0.11). Training data were collected from the ground representing approximately 10% of the study area covering the entire spectral gradient of the classes. Separate test data were collected for validation.

LISS-III MS classified image using conventional Bayesian classifier is shown in Figure 12(a). Assuming that there are six (fixed) number of representative endmembers (pure pixels), the entire image was modeled in terms of those spectral components, extracted using N-FINDR algorithm (Winter, 1999) from MODIS images. In the absence of pure pixels, alternative algorithms (Plaza et al., 2004) can be used for endmember extraction, which is a limitation of N-FINDR. It may be noted that some objects (for example buildings with concrete roofs, tiled roofs, asphalt, etc.) exhibit high degrees of spectral heterogeneity representing variable endmembers. This intra-class spectral variation with variable endmembers can be addressed through techniques discussed in (Bateson et al., 2000; Song, 2005; Foody and Doan, 2007).

Abundance values were estimated for each pixel through unmixing and used as prior probabilities in HBC to classify LISS-III MS data [Fig. 12 (b)]. Table 11 is the class statistics and Table 12 indicates the producer's and user's accuracies. The overall accuracy and Kappa for HBC (93.54%, 0.91) is higher than Bayesian classifier (87.55%, 0.85). For any particular class, if the reference data has more pixels with correct label, the producer's accuracy is higher and if the pixels with the incorrect label in classification result is less, its user's accuracy is higher (Mingguo et al., 2009). Bayesian classifier wrongly classified many pixels belonging to waste/barren/fallow as builtup. Forest class was over-estimated and plantation was under-estimated by Bayesian classifier. The only minority class in the study area is water bodies. Classified image using conventional Bayesian classifier had 1.08% (8854 ha) of water bodies in the study area. After the ground visit, we found that there are not many water bodies and the extents of most individuals were < 2000m2. Therefore, only a few water bodies that had spatial extent ≥ 62500 m2, could be used as endmembers. The minimum detected water class was 5% using unmixing of LS-HSR data. It may have happened that the prior probability was unlikely for this class while classifying the LISS-III MS data using HBC. The classified image obtained from HBC showed 0.81% (66.20 ha) of water bodies. A few pixels were wrongly classified using HBC, therefore, the producer's accuracy decreased from 90.91 (in conventional Bayesian classifier) to 88.18% (in HBC).

However, given the same set of training pixels for classification, the user's accuracy has increased from 88.89 to 97%. Producer's accuracy increased for agriculture (2.6%), builtup (4.3%), forest (7%), plantation (11.5%) and waste land (10.6%) and user's accuracy increased for agriculture (8%), built up (16.6%), forest (7.6%) and water bodies (8%) in HBC output. On the other hand, producer's accuracy decreased (2.7%) for water bodies and user's accuracy decreased (~0.7%) for plantation and waste land classes in agreement with the similar observations reported in Mingguo et al. (2009). HBC was intended to improve classification accuracies

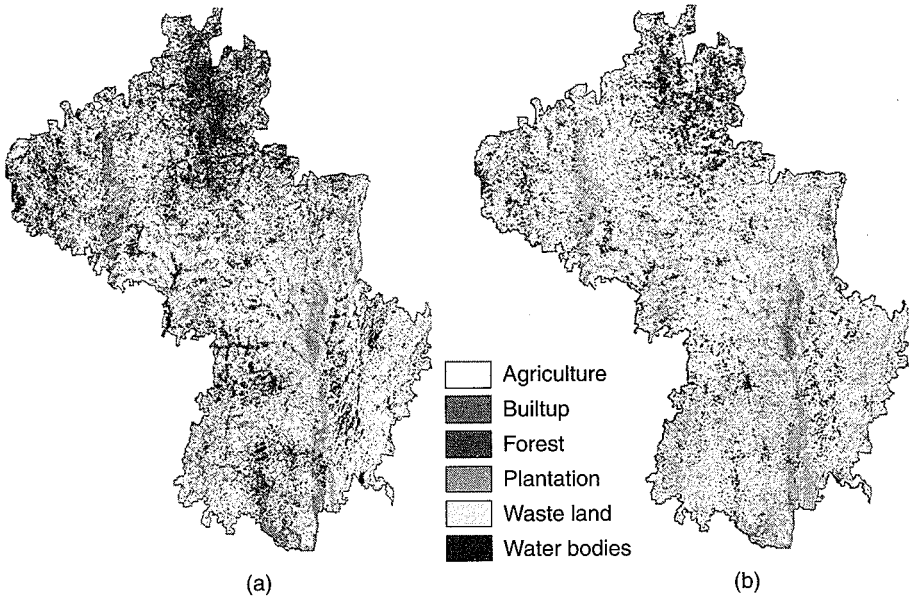

Fig. 12 LISS-III MS classified images: (a) Bayesian classifier, (b) HBC

Table 11 Class statistics from Bayesian and HBC for LISS-III data

Classifiers →	Bayesian classifier		HFC	
Class ↓	Ha	%	Ha	%
Agriculture	155,451	19.04	142931	17.51
Builtup	139,759	17.12	79280	9.71
Forest	93,241	11.42	55721	6.83
Plantation	89,493	10.96	176132	21.58
Waste land	329,473	40.36	355587	43.56
Water bodies	8854	1.08	66.20	0.81

by correctly classifying pixels which were likely to be misclassified by Bayesian classifier. Therefore a cross comparison of the two classified images located the pixels that were assigned different class labels at the same location. These wrongly classified pixels when validated with ground data revealed a 6% (~1742832 pixels) improvement in classification by HBC.

In the second experiment, IKONOS MS data (4 bands of 4 m, acquired on November 24, 2004) of 700 × 700 size and Landsat ETM+ data (6 bands excluding Thermal and Panchromatic, of 30m, acquired on November 22, 2004) of 100 × 100 dimension were co-registered (RMSE – 0.09). Landsat pixels were

Table 12 Accuracy assessment for LISS-III data

Classifiers →	Bayesian classifier		HBC			
Class ↓	PA*	UA*	PA*		UA*	
Agriculture	87.54	87.47	90.15	↑	95.56	↑
Builtup	85.11	81.68	89.39	↑	98.33	↑
Forest	85.71	88.73	92.61	↑	96.36	↑
Plantation	84.44	91.73	95.95	↑	91.03	↓
Waste land	88.03	90.37	98.67	↑	89.66	↓
Water bodies	90.91	88.89	88.18	↓	97.00	↑
Average	86.96	88.15	92.49	↑	94.66	↑

*PA – Producer's Accuracy; UA – User's Accuracy.

resampled to 28 m so that 49 IKONOS pixels would fit in 1 Landsat pixel. The scenes correspond to Bangalore city, India near the central business district having race course, bus stand, railway lines, parks, built up with concrete roofs, asbestos roofs, blue plastic roofs, coal tarred roads with flyovers and a few open areas (playground, walk ways, vacant land, etc.). Class proportions from Landsat image pixels were used as prior probabilities to classify the IKONOS MS images using HBC (Figure 13 (b)). The overall accuracy and Kappa for HBC (89.53%, 0.87) is higher than Bayesian classifier (80.46%, 0.69). Producer's accuracy increased by 17% for concrete, asbestos, blue plastic roof and open area and decreased by 7% for vegetation in HBC (Table 14). User's accuracy increased by ~5.5% for all the classes in HBC. Bayesian classifier wrongly classified and overestimated many pixels belonging to open area as asbestos roof (Table 13). Vegetation was overestimated by 9% using Bayesian classifier and open area was underestimated by ~15%. A cross comparison of the two classified images showed that 95733 pixels (0.33% of the study area) were differently classified by the two classifiers. Validation

Table 13 Class statistics for IKONOS data

Classifiers →	Bayesian classifier		HBC	
Class ↓	Ha	%	Ha	%
Concrete roof	346.67	44.34	356.18	45.56
Asbestos roof	47.99	7.41	6.79	0.87
Blue plastic roof	5.83	0.75	0.21	0.03
Vegetation	329.72	42.18	260.01	33.26
Open area	41.60	5.32	158.58	20.28

of these pixels showed an improvement of 9% by HBC, higher than accuracies reported in (Janssen and Middelkoop, 1992; Cetin et al., 1993; Strahler, 1980).

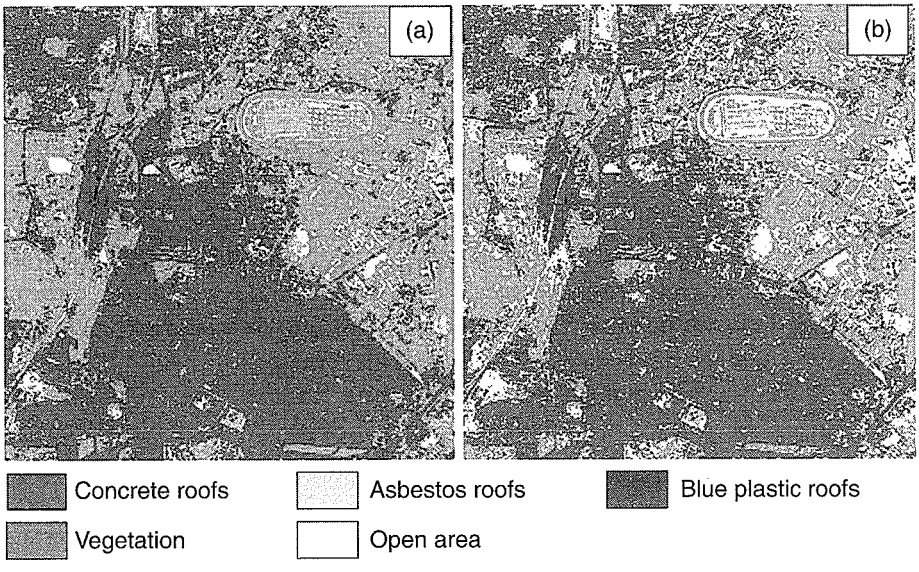

Fig. 13 IKONOS classified images: (a) Bayesian classifier, (b) HBC

Table 14 Accuracy assessment for IKONOS data

Classifiers →	Bayesian classifier		HBC			
Class ↓	PA	UA	PA		UA	
Concrete roofs	69.99	84.01	76.49	↑	93.89	↑
Asbestos roofs	84.77	87.77	91.89	↑	94.46	↑
Vegetation	94.21	87.55	87.24	↓	89.13	↑
Blue plastic roof	84.33	81.17	97.00	↑	85.60	↑
Open area	51.49	69.49	95.00	↑	74.22	↑
Average	76.96	81.99	89.52	↑	87.46	↑

7. CONCLUSION

In the first part of this chapter, seven advanced classifiers were tested and their performance analyses were carried out on multi-sensor data of varying spatial resolutions from IKONOS 4 m, IRS LISS-III MS 23.5 m, Landsat ETM+ 30 m to MODIS 250 m data. SMAP was best for IKONOS and ETM+ data classification,

KNN was superior to other techniques for LISS-III, and RF could classify MODIS data with maximum accuracy. In general, the performance of SMAP, KNN, SVM and RF are suitable for classification based exploratory data analysis using medium to high resolution sensor data. Due to the intrinsic scale of MODIS spatial resolution, spectral mixing cannot be avoided and ignored. In such cases, per-pixel classification do not provide much useful results and the low spatial resolution data need to be fused or merged with higher spatial resolution image (such as Panchromatic) through image fusion which increases the spatial as well as spectral property of data. However, improvements in spatial resolution of data do not alone suffice in producing accurate thematic maps. In such cases, classifying a complex landscape still remains a challenge and requires consideration of many other factors during classification and interpretation. RS data integrated with other ancillary geographical data such as elevation, aspect, temperature, texture, etc. can further improve the classification accuracy and may provide individual strategies for different landscape types such as highly mountainous regions (e.g. the Himalayas) to highly urbanized landscapes (such as Greater Bangalore City).

In the second part, a new Hybrid Bayesian Classifier was discussed. The major contribution of this technique lies in the fact that abundance estimates from low spatial-high spectral resolution data were utilized as prior probabilities to classify high spatial-low spectral resolution data using a Bayesian classifier, improving the overall accuracy by 6% and 9% with IRS LISS-III MS and IKONOS MS data respectively, as compared to conventional Bayesian classifier, demonstrating the robustness of the approach.

REFERENCES

1. Agarwal, R. and Srikant, R. 1994, Fast algorithms for mining association rules. In Proceedings of the 1994 International Conference VLDB, pp. 487–499, Santiago, Chile.

2. Atkinson, P. M., and Aplin, P., 2004, Spatial variation in land cover and choice of spatial resolution for remote sensing. International Journal of Remote Sensing, vol. 25, pp. 3687–3702.

3. Atkinson P. M., and Tatnall, A. R. L., 1997, Introduction: neural networks in remote sensing. International Journal of Remote Sensing, vol. 18(4), pp. 699–709.

4. Bateson, C. A., Asner, G. P., and Wessman, C. A., 2000, Endmember Bundles: A New Approach to Incorporating Endmember Variability into Spectral Mixture Analysis. IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 2, pp. 1083–1094.