Advanced Machine Learning Algorithms based Free and Open Source Packages for Landsat ETM+ Data Classification

Uttam Kumar[1], Anindita Dasgupta[2], Chiranjit Mukhopadhyay[3] and T.V. Ramachandra[4]

[1] Centre for Ecological Sciences, Department of Management Studies and Centre for Sustainable Technologies, Indian Institute of Science, Bangalore – 560012, India, e-mail: uttam@ces.iisc.ernet.in

[2] Centre for Ecological Sciences, Indian Institute of Science, Bangalore – 560012, India, e-mail: anindita_dasgupta@ces.iisc.ernet.in

[3] Department of Management Studies, Indian Institute of Science, Bangalore – 560012, India, e-mail: cm@mgmt.iisc.ernet.in

[4] Centre for Ecological Sciences, Centre for Sustainable Technologies, and Centre for *infrastructure*, Sustainable Transport and Urban Planning, Indian Institute of Science, Bangalore – 560012, India, email: cestvr@ces.iisc.ernet.in

***Abstract*—Detailed and accurate inventorying and mapping of land use (LU) at a regional/national scale is now possible with the availability of various medium spatial resolution sensors (such as Landsat, IRS LISS-III/IV, etc.). The LU information are derived using machine learning algorithms which are mainly dependent on the spectral properties of objects in the bands to assign them into a user defined class label. Designing a suitable image processing procedure is a prerequisite for successful classification of remotely sensed data into thematic information. Use of multiple features and selection of a suitable classification method are especially significant for improving classification accuracy. In this context, this paper reviews six advanced machine learning techniques such as Decision Tree, K-Nearest Neighbour, Neural Network (NN), Random Forest, Contextual Classification using sequential maximum a posteriori estimation (SMAP), and Support Vector Machine for Landsat ETM+ data classification using Free and Open Source (FOS) Packages along with the algorithmic descriptions. The ETM+ data classification results showed that SMAP classifier gave best performance with 89% overall accuracy and 0.8596 kappa followed by KNN with 87% overall accuracy and 0.8314 kappa while NN performed the last with 75% accuracy and 0.7142 kappa.**

Keywords-FOSS; machine learning; Landsat ETM+; classification

I. INTRODUCTION

With the development of remote sensing (RS) technology, space borne data have been widely used to classify land use (LU), permitting to update maps more frequently and nearly on a real-time basis [1]. Radiance/reflectance measurements obtained in various wavelength bands for each pixel provide spectral patterns that can be classified and correlated to different LU classes on the ground. Our ability to analyze RS data is important because it allows changes in the Earth's surface to be monitored as they occur [2]. By monitoring changes in the Earth's surface, a better understanding of the environmental control issues is possible. However, the unprecedented wealth of RS sensors and image-based geospatial information produce large volumes of data and result in large imagery-based data repositories [3]. Therefore, deriving LU information from these data is an important phase for the determination of land use/land cover information using machine learning algorithms.

The overall objective of classification is to assign all pixels in the image to particular classes or themes (e.g. water, forest, etc.). The resulting classified image represents a particular theme, and is essentially a thematic map of the original image [1]. In this context, machine learning algorithms tend to find hidden patterns, trends and relationships in data and classify them into user defined categories. Due to its broad applicability to many fields, machine learning has attracted tremendous attention from both researchers and practitioners.

In recent times, non-parametric methods, such as K-Nearest Neighbour, Neural Network, Random Forest, etc. have been recently practiced that have the advantage of not needing class density function estimation thereby obviating the training set size problem and the need to resolve multimodality [4-5]. In this paper, we perform a comparative analysis of six advanced machine learning algorithms viz. Decision Tree, K-Nearest Neighbour, Neural Network, Random Forest, Contextual Classification using sequential maximum a posteriori estimation (SMAP), and Support Vector Machine. These algorithms are evaluated and their performances are assessed on Landsat ETM+ data using Free and Open Source (FOS) Packages.

The paper is organised as follows. Section 2 describes the six machine learning algorithms along with a brief description of their FOS Packages, followed by data and study area in section 3. Section 4 presents the result and discussion and concluding remarks are given in section 5.

II. MACHINE LEARNING ALGORITHMS

A. 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 the analysis of training set. A tree structure is developed; a specific decision rule is implemented at each branch, which may involve one or more combination(s) of the attribute inputs. A new input vector then travels from the root node down through successive branches until it is placed in a specific class [6]. The 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 based on the frequency of each class at the node [7]. With minimum entropy, the stopping criterion is based on the amount of information gained by a rule (the gain ratio).

B. K-Nearest Neighbour (KNN)

The KNN algorithm [8] assumes that pixels close to each other in feature space are likely to belong to the same class. It bypasses density function estimation and goes directly to a decision rule. Several decision rules have been developed, including a direct majority vote from the nearest *k* neighbours in the feature space among the training samples, a distance-weighted result and a Bayesian version [9]. If **x** is an unknown pixel vector and suppose there are k_n neighbours labelled 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

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

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

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

The basic rule does not take the distance of each neighbour 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

$$m_n(\mathbf{x}) = \frac{\sum_{j=1}^{k_n} 1/d_{nj}}{\sum_{n=1}^{N}\sum_{j=1}^{k_n} 1/d_{nj}} \quad (2)$$

where d_{nj} is Euclidean distance. With *n* training samples, one needs to find the *k* nearest neighbours for every pixel in a large image. This means *n* spectral distances must be evaluated for each pixel. The algorithm is summarised as below. The variable "unknown" denotes the number of pixels whose class is unknown and the variable "wrong" denotes the number of pixels which have been wrongly classified.

set number of pixels = 0

set unknown = 0

set wrong = 0

For all the pixels in the test image

do

{

1. Get the feature vector of the pixel and increment number of pixels by 1.

2. Among all the feature vectors in the training set, find the sample feature vector which is nearest (nearest neighbour) to the feature vector of the pixel.

3. If the number of nearest neighbours is more than 1, then check whether the corresponding class labels of all the nearest sample feature vectors are the same. If the corresponding 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.

}

C. Neural network (NN)

NN classification overcomes the difficulties in conventional digital classification algorithms that use the spectral characteristics of the pixel in deciding the category of a pixel [7]. NN based Multi-Layer Perceptron (MLP) classification in RS 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 [5 and 10].

There are numerous algorithms to train the network for image classification. A comparative performance of the training algorithms for image classification by NN is presented in Zhou and Yang, (2010) [11]. The MLP in this work is trained using the error backpropagation algorithm. The main aspects here are: (i) the order of presentation of training samples should be randomised from epoch to epoch; and (ii) the momentum and learning rate parameters are typically adjusted (and usually decreased) as the number of training iterations increases. Back propagation algorithm for training the MLP is briefly stated in Kumar et al. (2011) [7].

D. Random Forest (RF)

RF are ensemble methods using tree-type classifiers $\{h(x,\Theta_k), k = 1,.....,\}$ where the $\{\Theta_k\}$ are i.i.d. (independent and identically distributed) random vectors and

$\mathbf{x}$ is the input pattern [12]. They are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It uses bagging to form an ensemble of classification tree [12]. RF is distinguished from other bagging approaches in that at each splitting node in the underlying classification tree, a random subset of the predictor variable is used as potential variable to define split. In training, it creates multiple CART (Classification and Regression Tree) trained on a bootstrapped sample of the original training data, and searches only across randomly selected subset of the input variables to determine a split for each node. It utilises Gini index of node impurity [13] to determine splits in the predictor variables. While classification, each tree casts a unit vote for the most popular class at input $\mathbf{x}$. The output of the classifier is determined by a majority vote of the trees that result in the greatest classification accuracy.

It is superior to many tree-based algorithms, because it lacks sensitivity to noise and does not overfit. The trees in RF are not pruned; therefore, the computational complexity is reduced. As a result, RF can handle high dimensional data, using a large number of trees in the ensemble. This combined with the fact that random selection of variables for a split seeks to minimise the correlation between the trees in the ensemble, results in error rates that have been compared to those of Adaboost, at the same time being much lighter in implementation. RF has also outperformed CART and similar boosting and bagging-based algorithm [14].

E. Contextual classification using sequential maximum a posteriori (SMAP) estimation

Spectral signatures are extracted from images based on training map by determining the parameters of a spectral class Gaussian mixture distribution model, which are used for subsequent segmentation (i.e. classification) of the multispectral (MS) images. The Gaussian mixture class describes the behaviour of an information class which contains pixels with a variety of distinct spectral characteristics. For example, forest, grasslands or urban areas are information classes that need to be separated in an image. However, each of these information 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 behaviour. Mixture classes improve segmentation performance by modelling each information class as a probabilistic mixture with a variety of subclasses. In order to identify the subclasses, clustering is first performed to estimate both the number of distinct subclasses in each class, and the spectral mean and covariance for each subclass. The number of subclasses is estimated using Rissanen's minimum description length (MDL) criteria. This criteria determines the number of subclasses which best describe the data. The approximate Maximum Likelihood estimates of the mean and covariance of the subclasses are computed using the expectation maximization (EM) algorithm.

SMAP improves segmentation accuracy by segmenting the image into regions rather than segmenting each pixel separately [15]. The algorithm exploits the fact that nearby pixels in an image are likely to have the same class and segments the image at various scales or resolutions using the coarse scale segmentations to guide the finer scale segmentations. In addition to reducing the number of misclassifications, the algorithm generally produces segmentations with larger connected regions of a fixed class. The amount of smoothing that is performed in segmentation is dependent on the behaviour of the data. If the data suggest that the nearby pixels often change class, then the algorithm adaptively reduces the amount of smoothing, ensuring that excessively large regions are not formed (http://wgbis.ces.iisc.ernet.in/grass/grass70/manuals/html70_user/i.smap.html).

F. Support Vector Machine (SVM)

SVM are supervised learning algorithms based on statistical learning theory and heuristics [1 and 16]. SVM map input vectors to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separates the data. The separating hyper plane maximises the distance between the two parallel hyper planes with an assumption that larger the margin between these parallel hyper planes, the better the generalisation error. The model produced by support vector classification only depends on a subset of the training data, because the cost function for building the model does not take into account training points that lie beyond the margin [1]. The success of SVM depends on the training process. The easiest way to train SVM is by using linearly separable classes.

G. Free and Open Source (FOS) Packages

The FOS Packages and model parameters to implement each algorithm (in Intel Pentium IV Desktop Computer, 3.00 GHz clock speed with 3.5 GB RAM and 1000 GB HD) are summarized next. For DT, rulesets were extracted using See5 (http://www.rulequest.com) with 25% global pruning. These rules were then used to classify Landsat ETM+ MS data. For KNN (algorithm was implemented in C programming language in Linux Platform), the number of nearest neighbour was kept 1 in feature space. In case of conflict, random allocation to LU class was done. NN based classification was implemented in C programming language, where a logistic function was used along with 4 hidden layer. Output activation threshold was set to 0.001, training momentum was set to 0.2, training RMS exit criteria was set to 0.1, training threshold contribution was 0.1, and the training rate was maintained at 0.2 to achieve the convergence. RF was implemented using a random forest package 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.

III. DATA AND STUDY AREA

A. *Study area* - A part of Central Western Ghats with gentle undulating hills, rising steeply from a narrow coastal strip bordering the Arabian sea to a plateau at an altitude of 500 m with occasional hills rising above 600 to 860 m (Fig. 1).

B. *Data* - Survey of India (SOI) Topographical Sheets (of 1:50000 and 1:250000 scales) were used to generate base layers. Field data were collected with pre-calibrated handheld GPS. Landsat ETM+ data (acquired on March 14, 2000) were downloaded from Global Land Cover Facility (http://www.landcover.org) Google Earth images (http://Earth.google.com) were used with the field data for validating classified outputs.

Figure 1. Location of the study area in Central Western Ghats (Google Earth Image).

IV. RESULTS AND DISCUSSION

Seven classifications were carried out with Landsat ETM+ bands 1 to 5 and band 7 of 2000 x 2000 size (acquired on March 14, 2000) using training data collected from field and validated using separate test data into agriculture, builtup, forest, plantation, wasteland and water bodies that are the six major categories in the forested and mountainous terrain of Uttara Kannada district in Central Western Ghats (Fig. 1). Figure 2 shows the classified images and the LU statistics are listed in Table I. Accuracy assessment was done by generating error matrix; producer's, user's, overall accuracies and kappa were computed (Table II). The highest two overall accuracies are highlighted in bold.

Landsat data having a spatial resolution of 30 m were classified most accurately using SMAP algorithm (89.03% overall accuracy as given in Table I). SMAP takes into account the intra class spectral variations and exploits spatial information among neighbouring pixels to improve classification results [17]. NN was difficult to train before it reached convergence as evident from the training RMS plot (not shown here). The area covered by this image has forested landscape, dominated by evergreen and semi-evergreen flora. 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 (Fig. 2).

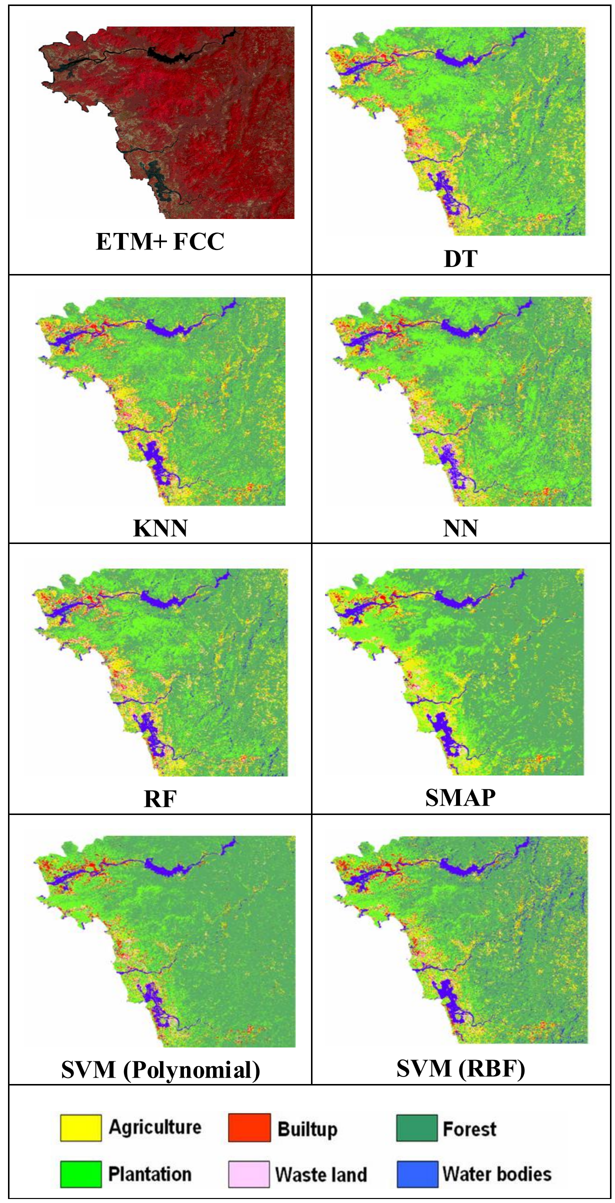

Figure 2. Classification of ETM+ Plus data through advanced classification algorithms.

Wasteland are often mixed with fallow land due to seasonal differences in crop practices, and also reflect similar to sand on the sea shores/sea beaches (Arabian sea on the west portion in the image), which were prominent in DT classification as depicted in Fig. 2. The above reported results are obtained with certain parameter settings and may not result in similar output when the parameters are altered or adjusted. At a regional scale, medium spatial resolutions such as Landsat TM/ETM+, Terra ASTER are most frequently used data. Eva et al., (2010) [18] demonstrates the usage of medium spatial resolution satellite imagery (Landsat-5 TM and SPOT-HRV) for monitoring forest areas from continental to territorial levels. Uncertainties involved in different stages of classification procedures influence classification accuracy, as well as the area estimation under different land use and land cover classes [19]. Understanding the relationships between the classification stages, identifying the probable factors that influence the accuracy and improving them are essential for successful image classification.

TABLE I. 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 II. ACCURACY ASSESSMENT FOR ETM+ CLASSIFIED DATA

Algorithm	Class	Producer's Accuracy (%)	User's Accuracy (%)	Overall Accuracy (%)	Kappa
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		

	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		

V. CONCLUSION

This work has shown the use of Free and Open Source Packages for advanced machine learning algorithms to classify Landsat ETM+ data. The analysis evaluated six algorithms such as Decision Tree, K-Nearest Neighbour, Neural Network (NN), Random Forest, Contextual Classification using sequential maximum a posteriori estimation (SMAP), and Support Vector Machine. SMAP classifier gave best performance with 89% overall accuracy followed by KNN with 87% overall accuracy. Neural Network did not performed well with lowest accuracy of 75%.

ACKNOWLEDGMENT

We are grateful to NRDMS Division, Ministry of Science and Technology (DST), Government of India and Indian Institute of Science, Bangalore for their financial and infrastructure support. We extend our sincere thanks to Global Land Cover Facility (GLCF), USA for providing the Landsat ETM+ imagery.

REFERENCES

[1] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *Int. J. Applied Earth Obs Geoinformation*, vol. 11, no. 5, pp. 352–359, 2009.

[2] R.J. Roiger and M.W. Geatz, Data Mining: A Tutorial-Based Primer, Addison Wesley, Boston. 2003.

[3] M. Tseng, S. Chen, G. Hwang and M. Shen, "A genetic algorithm rule-based approach for land-cover classification," *ISPRS J. Photogramm Remote Sens*, vol. 63, pp. 202–212, 2008.

[4] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2003.

[5] Y.V. Venkatesh and S. KumarRaja, "On the classification of multispectral satellite images using the multilayer perceptron," *Pattern Recognition*, vol. 36, pp. 2161–2175, 2003.

[6] S. Piramuthu, "Input Data for Decision Trees. Expert Systems with Applications," 2006, doi: 10.1016/j.eswa.2006.12.030.

[7] U. Kumar, N. Kerle, M. Punia and T.V. Ramachandra, "Mining Land Cover Information using Multilayer Perceptron and Decision Tree from MODIS data," *J. the Indian Society of Remote Sens*, vol. 38, no. 4, pp. 592–603, 2011.

[8] B.V. Dasarathy, (Editor), Nearest neighbor (NN) norms: NN pattern classification techniques, IEEE Computer Society Press, Los Alamitos, California, 1990.

[9] P.J. Hardin, "Parametric and Nearest-neighbor methods for hybrid classification: A comparison of pixel assignment accuracy," *Photogramm Eng Remote Sens*, vol. 60, pp. 1439–1448, 1994.

[10] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall International, Englewood Cli6s, NJ, 1999.

[11] L. Zhou and X. Yang, "Training Algorithm Performance for Image Classification by Neural Networks," *Photogramm Eng Remote Sens*, vol. 76, no. 8, pp. 945–951, 2010.

[12] L. Breiman, "Random Forests," *Machine Learning*, vol. 40, pp. 5–32, 2001.

[13] L. Breiman, J. Friedman, C.J. Stone and R.A. Olshen, Classification and Regression Trees, 3rd Ed., CRC Press, Boca Raton, Fl., pp. 372, 1998.

[14] P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, "Random Forests for land cover classification," *Pattern Recognition Letters*, vol. 27, pp. 294–300, 2006.

[15] C. Bouman and M.A. Shapiro, "Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans Image Process*, vol. 3, no. 2, pp. 162–177, 1994.

[16] X. Yang, "Parameterizing Support Vector Machines for Land Cover Classification," *Photogramm Eng Remote Sens*, vol. 77, no. 1, pp. 27–37, 2011.

[17] S. Magnussen, P. Boudewyn and M. Wulder, "Contextual classification of Landsat TM images to forest inventory cover types," *Int J. Remote Sens*, vol. 25, pp. 2421–2440, 2004.

[18] H. Eva, S. Carboni, F. Achard, N. Stach, L. Durieux, and J-F. Faure, "Monitoring forest areas from continental to territorial levels using a sample of medium spatial resolution satellite imagery," *ISPRS J. Photogramm Remote Sens*, vol. 65, no. 2, pp. 191-197, 2010.

[19] J.L. Dungan, Towards a comprehensive view of uncertainty in remote sensing analysis. In Foody, G.M. and Atkinson, P.M. (Eds), Uncertainty in Remote Sensing and GIS, Chichester: John Wiley & Sons, pp. 23-35, 2002.

AUTHORS PROFILE

Uttam Kumar has an Engineering Degree in Computer Science from VTU, Belgaum, India., Master's Degree in Geoinformation Science from University of Twente, The Netherlands and a Ph.D. from the Indian Institute of Science, Bangalore. His areas of research are developing algorithms for multi-satellite sensor spatial temporal data analysis. His research interests are Pattern recognition, Remote sensing, Data mining and Image processing using Free and Open Source Software (FOSS).

Anindita Dasgupta is with the Centre for Ecological Sciences, Indian Institute of Science, Bangalore. She has a Master's Degree in BioTechnology from Bangalore University, India. Her research interests include Remote Sensing, Image Processing, Geographical Information Systems, Spatial Data Analysis and Urban Sprawl.

Chiranjit Mukhopadhyay earned a Ph.D. in Statistics from University of Missouri, Columbia in 1992, after obtaining M-Stat and B-Stat from Indian Statistical Institute, Kolkata. Currently he is Associate Professor of Statistics in the Department of Management Studies, Indian Institute of Science, Bangalore. He has served in the faculty of several Universities around the globe like The Ohio State University, Case Western Reserve University, Bilkent University etc. His current research interests include Bayesian Statistics, Reliability Theory, Statistical Quality Control, Empirical Finance and Financial Econometrics.

Ramachandra T. V. obtained his Ph.D. from Indian Institute of Science (IISc), Bangalore. Currently he is the Co-ordinator of Energy & Wetlands Research Group at the Centre for Ecological Sciences (CES), Associate Faculty at Centre for Sustainable Technologies (CST) and Centre for *infrstructure*, Sustainable Transport and Urban Planning (C*i*STUP) at IISc, Bangalore. His area of research includes remote sensing, digital image processing, urban sprawl: pattern recognition, modelling, energy systems, renewable systems, energy planning, energy conservation, environmental engineering education, etc. He is member of many national and internationally recognized professional bodies and is a recipient of the prestigious Satish Dhawan Young Engineer Award, 2007 of Karnataka State Government, India.