

Decision tree approach for classification of remotely sensed satellite data using open source support

RICHA SHARMA, ANIRUDDHA GHOSH and P K JOSHI*

Department of Natural Resources, TERI University, New Delhi 110 070, India.

**Corresponding author. e-mail: pkjoshi27@hotmail.com*

In this study, an attempt has been made to develop a decision tree classification (DTC) algorithm for classification of remotely sensed satellite data (Landsat TM) using open source support. The decision tree is constructed by recursively partitioning the spectral distribution of the training dataset using WEKA, open source data mining software. The classified image is compared with the image classified using classical ISODATA clustering and Maximum Likelihood Classifier (MLC) algorithms. Classification result based on DTC method provided better visual depiction than results produced by ISODATA clustering or by MLC algorithms. The overall accuracy was found to be 90% ($\kappa = 0.88$) using the DTC, 76.67% ($\kappa = 0.72$) using the Maximum Likelihood and 57.5% ($\kappa = 0.49$) using ISODATA clustering method. Based on the overall accuracy and kappa statistics, DTC was found to be more preferred classification approach than others.

1. Introduction

Research using remotely sensed satellite data has attracted attention on image classification because classification results are basis for interpretation, analysis and modelling for various environmental and socio-economic applications (Jensen and Cowen 1999). Researchers have made efforts in developing and applying various classification approaches and techniques (Lu and Weng 2007) with the objectives of identifying representative spectral values (Shaw and Burke 2003; Jensen 2005), mapping more number of classes like (Anderson *et al.* 1976; Gregorio and Jansen 2000) representing real-time land use land cover (LULC) map and improving classification accuracy (Aguirre-Gutiérrez *et al.* 2012) and others. Still, classification of remotely sensed satellite data into a thematic map remains a challenge. This is attributed to many

factors including geometry of satellite and source of illumination, terrain characteristics (Dorren *et al.* 2003; Balthazar *et al.* 2012), time and season of image capture (Moody and Strahler 1994) and many others.

Satellite data classification is a method by which labels or class identifiers are attached to individual pixels on the basis of their spectral characteristics in various bands. Traditional classification procedures like maximum likelihood classifier (MLC) are generally based on statistical parameters such as mean and standard deviation. These classifiers are generally characterized by having an explicit underlying probability model, which provides a probability measure of any pixel in each class rather than simply a class value. Some of the other common classification algorithm include the nearest neighbours (NN or kNN), minimum distance to means, ISODATA (Mather 2004; Lillesand

Keywords. Remote sensing; image classification; decision tree classifier (DTC); maximum likelihood classifier (MLC); ISODATA.

et al. 2007). The performance of this type of classifier depends on how well the data match the pre-defined model (Jensen). These models generally fail to give the desired output if the data are complex. In order to overcome such problems, more advanced classifiers like artificial neural networks (Pu *et al.* 2008), fuzzy classifier (Feitosa *et al.* 2009; Thapa and Murayama 2009), image segmentation (Gamanya *et al.* 2007, 2009), expert classification (Stefanov *et al.* 2001; Wentz *et al.* 2008), support vector machines (Carräa *et al.* 2008) and many others are increasingly being used. Most of these techniques do not make any *a priori* assumption about the data distribution, hence are essentially non-parametric in nature. Decision Tree classification (DTC) is one such technique which is very effective and useful for the remote sensing community for LULC classification (Punia *et al.* 2011). This is intuitive, simple, flexible, and efficient in computing (Friedman 1977) non-normal, non-homogenous and noisy data as well as non-linear relation between features and classes, missing value, and both numeric and categorical inputs (Quinlan 1993). This has led to increased acceptance (Quinlan 1993; Pal and Mather 2003).

Otukei and Blaschke (2010) advocated the use of open source data mining software, such as WEKA, for creating the decision boundaries instead of applying expert human knowledge. They also argued that use of automated data mining system can overcome the problem of disagreement between the decision boundaries proposed by various experts. In this study, open source remote sensing datasets and data mining software support for LULC classification have been explored. The main objectives pursued are: (1) to develop a DTC algorithm for remotely sensed satellite data using open source datasets, (2) to compare the performance of the DTC, ISODATA and MLC using standard measures of accuracy assessment.

2. Decision tree classifier

The idea of using decision trees to identify and classify objects was first reported by Hunt *et al.* (1996). It was inspired by Automatic Interaction Detection (AID) program (Morgan and Sonquist 1963) and successfully followed by the *Chi*-squared Automatic Interaction Detector (CHAID) (Morgan and Messenger 1973). Later, induction of Decision Tree (ID3) software package (Quinlan 1975), Classification and Regression Trees (CART) (Breiman *et al.* 1984), C4.5 (Quinlan 1993), and Oblique Classifier (OC1) package (Murthy *et al.* 1994) were developed. These are non-parametric classifiers.

Decision tree classification techniques have been used successfully for a wide range of classification problems, but only recently been tested in detail by the remote sensing community (Friedl and Brodley 1997; Pal and Mather 2003; Kandrika and Roy 2008; Wen *et al.* 2008). Several studies have compared DTC methods with other classifiers. Otukei and Blaschke (2010) compared decision tree, maximum likelihood and support vector machine based techniques for land cover change assessment using Landsat TM and ETM+ data and found decision tree based methods performed better than others. Punia *et al.* (2011) used C5.0 based decision tree classifiers to classify IRS-P6 AWiFS data and reported very high accuracy. Duro *et al.* (2012) compared decision tree, support vector and random forest methods for the classification of agricultural landscapes using SPOT-5 HRG imagery in both pixel and object oriented domain and found for the specific case study, all the algorithms equally performed. Many of the studies concentrated on improving the accuracy of DTC using advanced algorithms and incorporating ancillary datasets. Evrendilek and Gulbeyaz (2011) used boosted decision tree classifications for mapping LULC from MODIS satellite images, climate and topographic data. Possibility of incorporating various types of data sources under a single classifier framework has proved to be advantageous in using DTC. Researchers have used DTC for continental to global land cover classification problems with improved accuracy (Friedl *et al.* 1999; Hansen *et al.* 2000).

The construction of decision tree requires supervised training; therefore it is necessary to have a training dataset consisting of response and explanatory variables. In classification problems involving remote sensing dataset, the response variables are generally LULC classes and explanatory variables are spectral bands or information derived from these. The classification structure defined by a decision tree is estimated from training data using a statistical procedure. The 'tree' is made of a root node, internal nodes and leaves. Nodes are where trees branch or split the dataset; terminal nodes are called leaves which contain most homogeneous classes. If in a training set T , there is k number of classes (C) and a total of $|T|$ cases, the expected information from such a system is,

$$\text{info}(T) = - \sum_{j=1}^K \frac{\text{freq}(C_j, T)}{|T|} \log_2 \left(\frac{\text{freq}(C_j, T)}{|T|} \right)$$

where, $\frac{\text{freq}(C_j, T)}{|T|}$ is the probability of occurrence of class C_j in training set T . If we partition the training set T in accordance with any response variable X (e.g., NDVI), there may be ' n ' number of cases.

The expected information after this partition is weighted sum over the subsets as:

$$\text{infox}(T) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \text{info}(T_i).$$

The information gain by portioning training set following variable X (NDVI) is

$$\text{gain}(T) = \text{info}(T) - \text{infox}(T).$$

The gain criterion selects a test (on the basis of the explanatory variables) to maximize the information gain. But this has a strong bias, favouring the tests with many outcomes. This is rectified by introducing gain ratio (X), defined as:

$$\text{gain ratio}(X) = \text{gain}(X) / \text{split info}(X)$$

where, $\text{split info}(X)$ is the potential information generated by dividing T into n subsets, which is not relevant to classification and is given by:

$$\text{split info}(X) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \log_2 \left(\frac{|T_i|}{|T|} \right).$$

Gain ratio represents the proportion of useful information for classification created due to the split. The experiment (on the basis of explanatory variables) which maximizes the ratio, subject to the constraint of large information gain is selected. Training samples (root node) are partitioned following the mentioned condition to create branches. At some node, sample distribution in each group becomes uniform, i.e., only one type of LULC class is present in each group and hence information content is largest. These are defined as 'leaves' (pure classes). Additional details can be found in Quinlan (1993). In this study, an open source Java implementation of the C4.5 algorithm in the WEKA data mining tool, known as J48 was used.

3. Materials and methods

3.1 Study area

Surat city, the commercial capital city of Gujarat (India) serves as the administrative capital of the

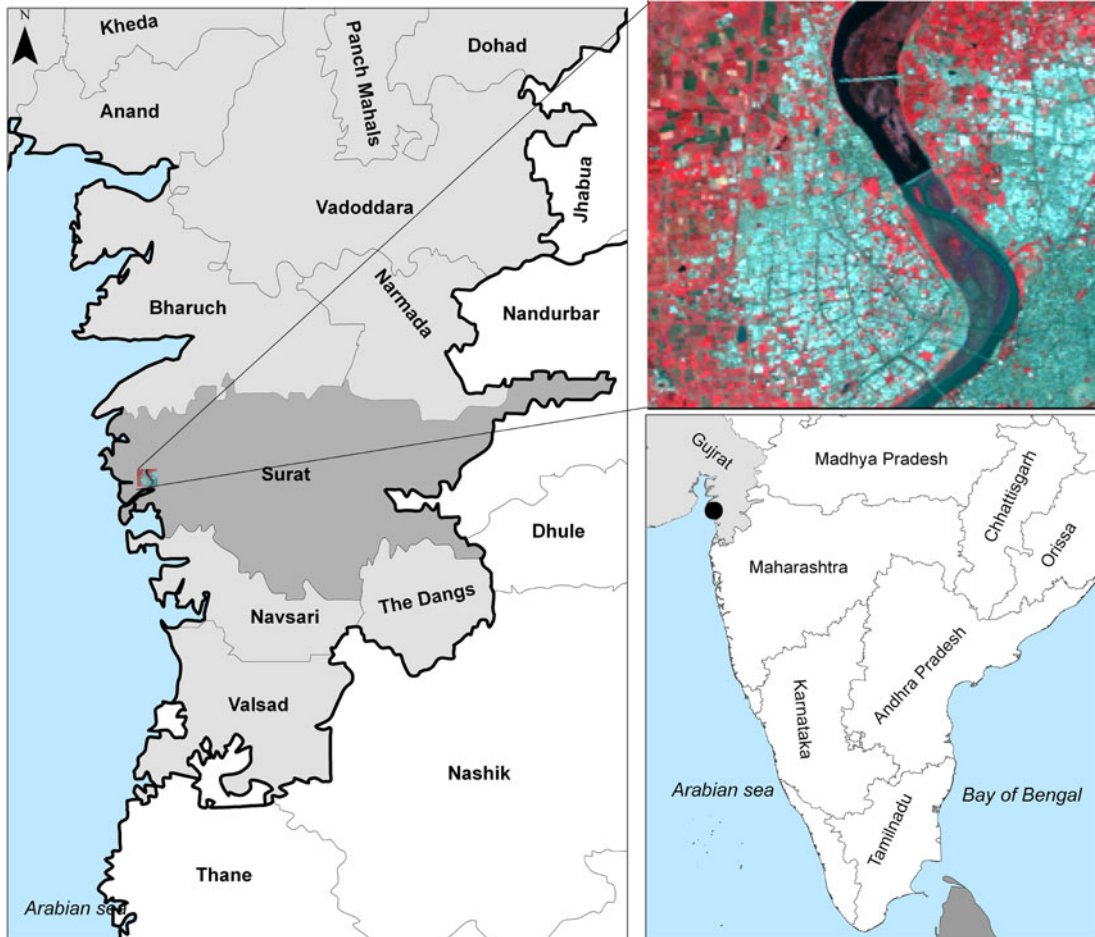


Figure 1. Location of the study area.

Surat district. The city is also called the ‘Diamond City’, and is situated on the banks of Tapi river which perennially flows northeast to south-west finally joining the Arabian Sea. The city is located at 21.25°N and 72.87°E (figure 1). The region experiences hot summers with temperatures ranging from 38° to 45°C. Winters are mild with temperature in the month of January varying between 10° and 15.5°C. The average annual rainfall is 1143 mm (www.suratmunicipal.gov.in). Being a commercial hub, the region has witnessed tremendous urbanization over past few decades. It ranks 36th in the list of world’s largest cities with population over 2.5 million (population density of 1376 persons/km²) (Census of India 2011) and third cleanest metropolitan region in India. Surat has immense business and job opportunities that trigger high immigration rates to the city resulting in increased population. The study area extends spatially from 21.29°–21.12°N to 72.74°–72.94°E, spanning across 386.28 km² covering the city (city limits of 326 km²) and the peripheral agricultural fields along with part of the Tapi river. The study area mainly falls in the Chorasi taluk of Surat district but also covers parts of adjoining taluks of Olpad in north, Kamrej in east and Palsana in south-east.

3.2 Remote sensing data

Landsat5 Thematic Mapper (TM) sensor system was launched on March 1, 1984. It operates in four spectral bands in VNIR region; Band1 (blue, 0.45–0.52 μm), Band2 (green, 0.52–0.60 μm), Band3 (red, 0.63–0.69 μm), Band4 (NIR, 0.76–0.90 μm) and two bands in MIR (Band5, 1.55–1.75 μm and Band7, 2.08–2.35 μm) and one band is thermal (Band6, 10.4–12.5 μm). With a GIFOV of 30 m \times 30 m for bands 1 through 5 and 7 and 120 m \times 120 m for band 6, Landsat TM provides temporal resolution of 16 days. The USGS offers nearly cloud free ortho-rectified data in GeoTIFF format. TM satellite image (Path/Row: 148/45) dated

23 October 2009 (figures 1 and 4a) was used for the study. The geometrically and radiometrically corrected image rectified to a common Universe Transverse Mercator (UTM) was procured from the USGS Earth Resource Observation Systems (EROS) Data Center.

3.3 Classification

Classification of remotely sensed data is one of the primary steps for information extraction. This is done using various classification techniques that could be parametric, or non-parametric, hard or soft. These techniques are used to assign one of the various classes to each pixel based on its spectral response. In this paper three classification methods were used; ISODATA (Iterative Self-Organizing Data Analysis) Clustering, MLC and DTC to map out six classes based on classification scheme (table 1) using each technique.

The satellite data was clustered into 50 classes using ISODATA with six iterations and 0.95 convergence threshold. These clusters were then assigned to one of the six land use categories identified and were merged to achieve unsupervised classification.

Another classification technique applied was supervised classification using MLC. This method calculates the probability of a pixel belonging to each of the classes under consideration. Based on the maximum probability, the pixel is then assigned a class. The probability that x will fall under a particular class w , is given by:

$$p(x|w_i) = \frac{1}{(\sigma_i \sqrt{2\pi})} \exp[-(x - \mu_i)^2 / (2\sigma_i^2)].$$

The decision tree classification was performed using WEKA. The image was converted into ASCII format and was then used to perform DT classification. An important step of DT classification is rules, generation which is explained in the following section. Decision rule set was generated using

Table 1. *Classification scheme.*

Land use class	Description
Built-up (dense)	Area with high density of built-up land use.
Built-up (sparse)	Areas with relatively less density of built-up, could be interspersed by vegetation or open area or other land use.
Agriculture	Land under cropping, also includes plantation areas.
Fallow	Agricultural land but is currently fallow due to harvest.
Water (deep)	Water bodies that are natural like river or ponds; or artificial water storage tanks etc.
Water (shallow)	Land that is under water but is shallow or highly turbid, also river beds that are submerged under shallow water.

training sets in WEKA J48 classifier. WEKA is an open source data mining software that has a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or can be called from a Java code. It has various tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The J48 decision tree algorithm available with WEKA has been used for training the Landsat TM dataset (Quinlan 1993). The output rule sets along with the trial classification results were examined for their confidence levels and accuracies. Based on these results, the training sites were modified wherever necessary till reliable training sets are obtained, and good classification accuracies (based on Kappa statistics and overall accuracy) were achieved. The rule set with highest accuracy was used to classify entire dataset in WEKA using J48 classifier. The image was classified into six land use classes using training from signature dataset consisting of 644 training pixels of which 8% were for deep water, 9% were for shallow water, 11% were for sparse and dense built-up each, 19% for agriculture and rest 42% for fallow land.

3.3.1 Classification performance analysis

Classification performance is dependent upon four crucial factors; class separability, training sample size, dimensionality and classifier type. Class separability was studied using the Transform Divergence (TD) test. The test works by assigning exponentially decreasing weight to increasing distance between the classes. The results range from 0 to 2000. Classes exhibit good separability, if TD test value is greater than 1900, if value is between 1700 and 1900 the separation is fair and it is poor for values below 1700. Additionally feature space plot was analysed to evaluate the training sets.

Accuracy assessment measures the classification performance based on different classifiers. Independent ground sample points, uniformly distributed throughout the study area, were collected from satellite data and fine resolution Google

Earth images. Statistically valid sampling strategy was adopted to assess commission, omission and overall accuracy (Stehman 1996). A set of LULC information collected in this method was used for accuracy assessment. Cover type information of these locations was compared with the classified map.

4. Results and discussion

4.1 Class separability

Class separability was performed through TD test and feature space plot. TD test was performed in ERDAS and the results show that classes are highly separable with each pair showing TD test value of 2000 which is the maximum value (table 2).

The feature space plot (figure 2) was constructed to further analyse the separability of classes. The feature plot further complemented the observation that most of the classes could be separated from each other. But feature plot also demonstrated that

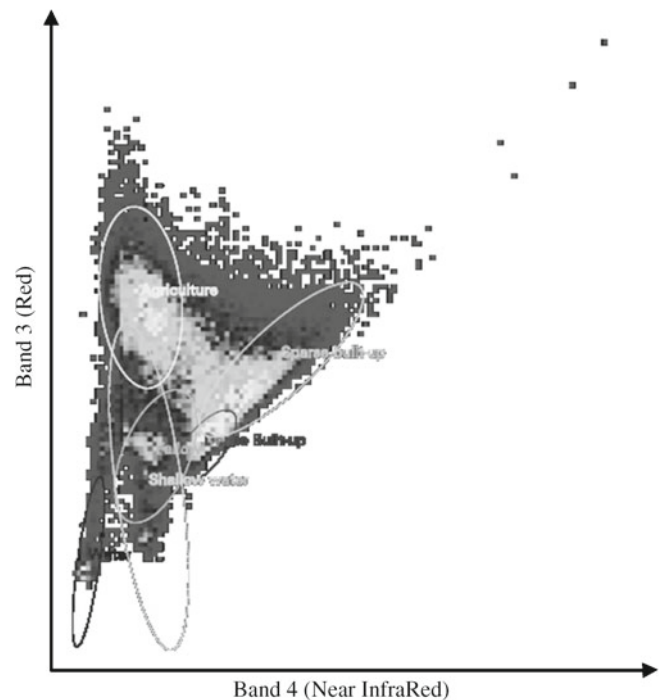


Figure 2. Feature space plot NIR vs. red band.

Table 2. Transformed divergence test results among different class pairs.

	Sparse built-up	Agriculture	Fallow	Deep water	Shallow water
Dense built-up	2000	2000	2000	2000	2000
Sparse built-up		2000	2000	2000	2000
Agriculture			2000	2000	2000
Fallow				2000	2000
Deep water					2000

major overlap exists for shallow water and fallow class, and some overlap was observed between sparse and dense built-up classes.

4.2 Decision tree

Decision tree generated using WEKA J48 classification algorithm is shown in figure 3. The important parameters of the algorithm along with their status in this study are described in table 3. The pruned tree is modelled in 0.17 seconds with

97.82% correctly classified instances. The decision tree classifies water (deep and shallow), built-up (sparse and dense), agricultural area and fallow lands.

B4 (NIR) is the band used for initial binary splitting of the image, which is subsequently supported by B2 (green). In the first node, B4 (NIR) and B2 (green) are used to segregate agriculture and dense built-up from other classes. The second node uses B1 (blue), B4 (NIR), and B7 (mid-IR) to segregate water (deep), fallow and sparse built-up (or

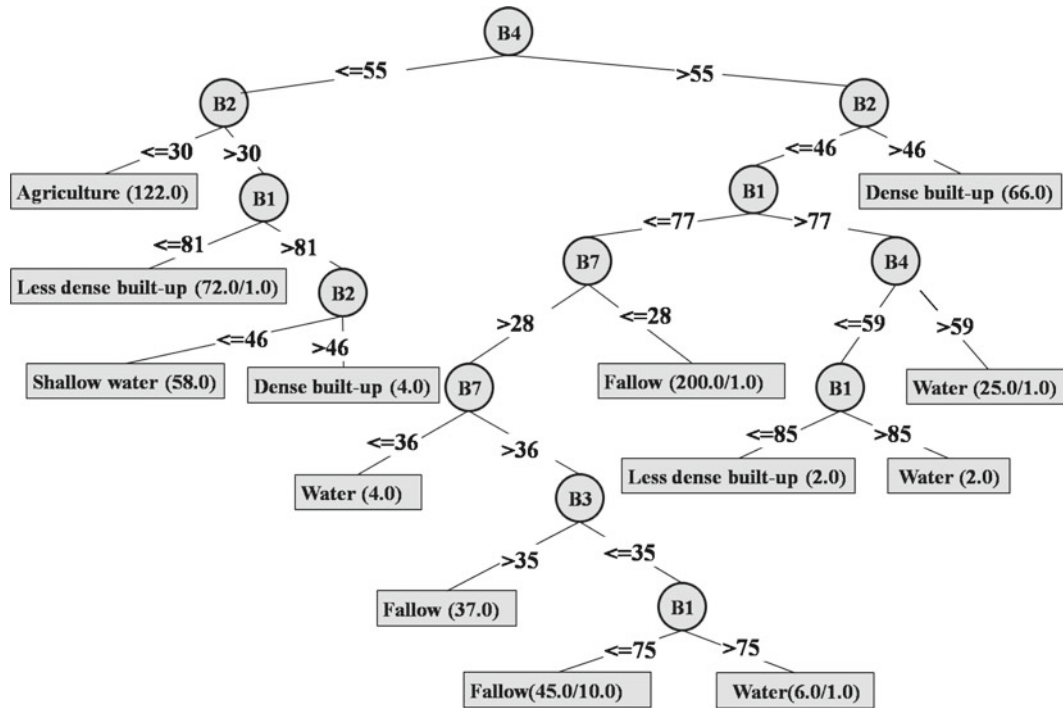


Figure 3. The pruned decision tree generated in WEKA.

Table 3. Explanation of WEKA J48 parameters.

Parameter	Description	Status
Binary splits	Whether to use binary splits at each node to build the decision tree	Used
Confidence factor	The confidence factor used for pruning; smaller confidence factor implies more pruning; the default value is 0.25	0.25
Number of objects	Minimum number of instances (sample) per leaf; if less number of samples are present in one leaf than the assigned value, the leaf will not be considered as a class	5
Number of folds	Determines the amount of data used for reduced error pruning; one fold is used for pruning, rest is used for growing the tree	Value set to 6; 1 fold is used for pruning and 5 fold is used for growing the tree
Size of tree	Number of nodes	25
Number of leaves	This is same as number of rules	13

less dense). B1 (blue), B2 (green) and B4 (NIR) are found useful in extracting shallow water class. B5 (mid-IR) was not found significant to differentiate among the classes.

4.3 Evaluation of training sets

Evaluation of sample pixels is crucial as training data is found to significantly influence classification performance. Accuracy of decision rules for instance is influenced by accuracy of training site characteristics. Too few or too many training pixels for a particular class may result in over- or under-estimation of that particular class in final outputs. Differences in training strategies can result in differential classification accuracies and results (Chen and Stow 2002). Thus, sample data evaluation is done using WEKA.

Tables 4–6 provide evaluation of the training data for the rule generation and details of rules generated and rules used for classification. Sampling accuracy is 97.82% with a kappa of 0.97 (table 4).

Table 5 is the confusion matrix for classified sample data, representing a measure of accuracy for each class. All classes except deep water and fallow show 100% accuracy. Producer accuracy for fallow is 99.3% and that for deep water is 76.1%. While the training data for two classes show user accuracy of 96% and 95.6% respectively.

Since, overall sampling accuracy has disadvantage of not being sensitive to class distribution, ROC area or AUC (Area Under Curve) is often cited as more important criteria to judge the accuracy of training datasets. ROC area for this study for each class is observed to be high ranging from 0.98 to 1 (table 6).

4.4 Classification results

Classification results of three different classification methods along with a false colour composite (FCC) of satellite data is presented in figure 4. A preliminary qualitative assessment shows that MLC and DTC exhibit similar distribution pattern of shallow water and fallow class. In MLC and ISODATA, many dense built-up pixels have been categorised as sparse built-up. The built-up area represented by bridge has been accurately mapped in DTC and ISODATA. Overall DTC classified map shows best classification results by differentiating between various land-use parcels.

Table 4. Evaluation parameters for classification.

Parameter	Value
Correctly classified instances	97.82%
Incorrectly classified instances	2.18%
Kappa statistic	0.97
Mean absolute error	0.01
Root mean squared error	0.07
Relative absolute error	4.82%
Root relative squared error	21.95%
Coverage of cases (0.95 level)	99.53%
Mean rel. region size (0.95 level)	17.98

Table 5. Confusion matrix for sampling accuracy.

	Built-up (dense)	Built-up (sparse)	Agriculture	Fallow	Water (deep)	Water (shallow)
Built-up (dense)	70	0	0	0	0	0
Built-up (sparse)	0	73	0	0	0	0
Agriculture	0	0	122	0	0	0
Fallow	0	0	0	271	2	0
Water (deep)	0	0	0	11	35	0
Water (shallow)	0	0	0	0	0	58

Table 6. Sampling accuracy details.

Class	TP rate	FP rate	Precision	Recall	F-measure	MCC	ROC area	PRC area
Dense built-up	1	0	1	1	1	1	1	1
Less dense built-up	1	0.002	0.986	1	0.993	0.992	0.999	0.987
Agriculture	1	0	1	1	1	1	1	1
Fallow	0.993	0.03	0.961	0.993	0.977	0.959	0.997	0.991
Water	0.745	0.003	0.946	0.745	0.833	0.829	0.982	0.848
Shallow water	1	0	1	1	1	1	1	1
Weighted average	0.978	0.013	0.978	0.978	0.977	0.969	0.997	0.984

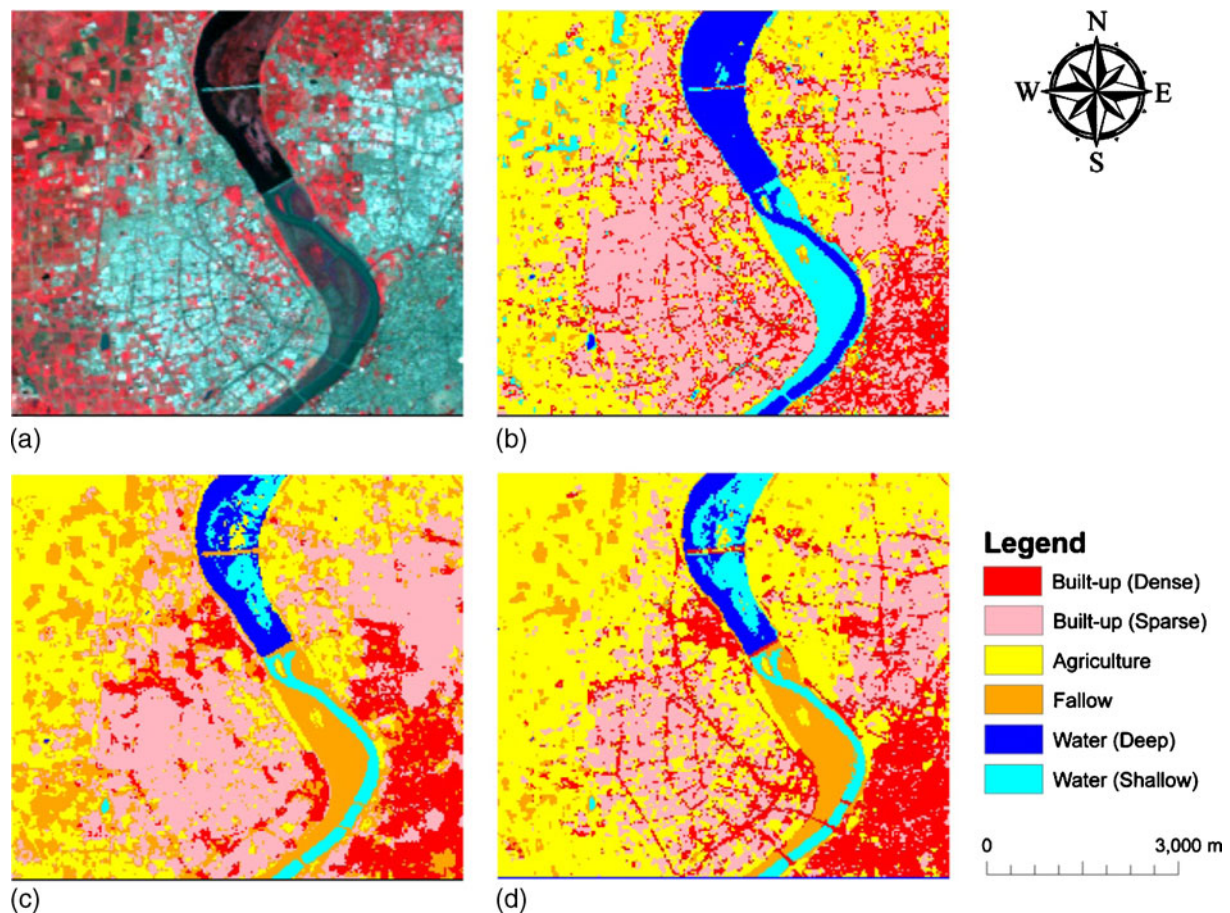


Figure 4. (a) Landsat TM standard false colour composite, (b) LULC map using unsupervised classification, (c) LULC map using supervised classification and (d) LULC map using decision tree classification.

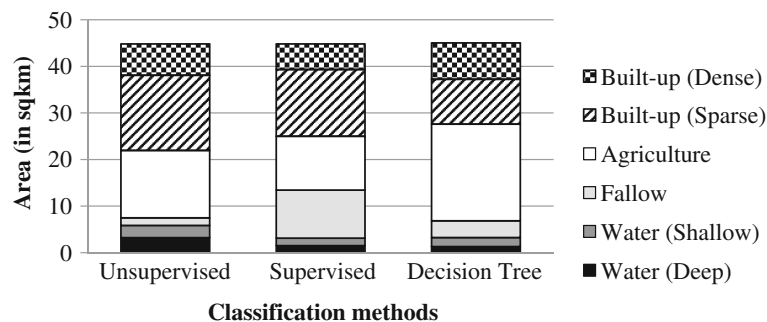


Figure 5. Area distribution in different classification schemes.

A quantitative comparison for distribution of different land use classes through different classification methods shows that both shallow (2.63 km^2) and deep water (3.25 km^2) and sparse built-up (16.15) has been overestimated in ISODATA classification as against MLC (1.63 , 1.5 and 14.3 km^2 , respectively) and DTC (1.94 , 1.31 and 9.62 km^2 , respectively). Maximum agriculture

(20.76 km^2) has been mapped by DTC method, while MLC estimates maximum fallow class (10.31 km^2) as compared to other two classifiers. Unsupervised ISODATA shows minimum coverage of fallow (1.63 km^2) among the three algorithms. Area distribution in different classes derived using classification schemes is shown in figure 5.

Table 7. Classification accuracies for three classification algorithms.

Class	ISODATA		MLC		DTC	
	PA	UA	PA	UA	PA	UA
Built-up (dense)	47.22	60.71	61.11	84.62	97.22	87.5
Built-up (sparse)	86.36	62.3	77.27	64.15	81.82	90
Agriculture	83.87	74.29	74.19	79.31	96.77	75
Fallow	4.17	100	81.25	72.22	79.17	95
Water (shallow)	2.33	3.23	69.77	96.77	93.02	100
Water (deep)	100	46.34	94.74	78.26	97.37	92.5
	OA=50.8	K=0.41	OA=76.67	K=0.72	OA=90	K=0.88

PA: Producer's accuracy, UA: User's accuracy, OA: Overall accuracy, K: Kappa statistics.

4.5 Accuracy assessment

For quantitative analysis of classification performance by three classifiers, accuracy assessment is performed. A confusion matrix was generated after overlaying the reference locations on the classified map. Classification using the DTC gave best overall accuracy of 90% and kappa as 0.88. It is followed by supervised classification giving overall accuracy of 76.67% and kappa of 0.7186. Overall accuracy for classification using ISODATA 50 clusters merged into eight classes gave overall accuracy as 50.83% and overall kappa as 0.4134. Classification accuracy for ISODATA ranges from 2.33% (PA for shallow water) to 100% (PA for deep water and UA for fallow). MLC accuracy varies from 61.1% (PA for dense built-up) to 96.8% (UA for shallow water). In comparison to other two classifiers, DTC exhibit highest accuracy range of 75% (UA for agriculture) to 100% (UA for shallow water) (table 7).

5. Conclusion

The study demonstrated use of open source remote sensing datasets and data mining software support for LULC classification. Open source data mining software WEKA was used to develop a DTC algorithm for remotely sensed satellite data. DTC outcomes were compared with outputs of ISODATA clustering and MLC classifier. The comparison of the classification algorithms shows DTC to be superior to the other two. Work by Otukey and Blaschke (2010) also shows that DTC performed better than both MLC and SVMs for similar kind of studies using Landsat data.

The DTC produced consistently higher classification accuracies than ISODATA and MLC. Better performance of the DTC is attributed to its ability to adapt to the noisy and non-linear relations often observed between the land use classes in the urban dominated landscapes and irregular

patterns. Being a non-parametric, it can easily ignore distribution properties of input data. A unique strength of DTC lies in their flexibility and simplicity for partitioning datasets as a function of the input feature space. At the same time, it employs differentiation among the linear feature interspersed with unlike classes while defining boundaries between classes. Examination of the class-specific classification accuracies provided by each of the candidate classifiers illustrates several important points. Most importantly, decision tree algorithms often optimize for maximum overall classification accuracy at the expense of smaller classes. However, this also depends on the cell or pixel size of the satellite data.

Open source data mining software use attributes of a pixel in different bands to construct a decision tree. These softwares have additional advantage of their ease to adapt and implement the classification algorithms. This is not possible using decision tree tools in proprietary software packages. Another added advantage of open source data mining software is faster processing and advantage of flexible customization based on background knowledge. These are also easy to procure and give satisfactory results for majority of applications.

The study has witnessed some of the limitations due to the nature of software and the data itself. WEKA has a huge limitation in handling large datasets. Thus, the methodology could be implemented only for smaller area. In addition to this, the spatial resolution of the data is not sufficient for analyzing finer details. The study also lacks in ground data collection.

The study attempts to highlight the advantages of DTC over conventional classification methods. DTC requires much less effort on part of user to prepare data for building decision trees. Being non-parametric in nature, the classifier is not sensitive to outliers as splitting is based upon proportion of samples within split ranges and not on absolute values. Additionally, DTC is much easier to implement and understand.

Acknowledgements

RS and PKJ acknowledge Department of Science and Technology (DST), Ministry of Science and Technology, Government of India. AG acknowledges Council of Scientific and Industrial Research (CSIR), Government of India for support. The experimental and computational facilities of TERI University have been used; support to this research is greatly acknowledged by authors. Authors also acknowledge anonymous reviewers for constructive comments and suggestions to improve quality of manuscript.

References

- Aguirre-Gutiérrez J, Seijmonsbergen A C and Duivenvoorden J F 2012 Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico; *Appl. Geogr.* **34** 29–37.
- Anderson James R, Hardy E E, Roach J T and Witmer R E 1976 A Land Use and Land Cover Classification System for use with Remote Sensor Data; USGS Professional Paper 964. A revision of the land use classification system as presented in the USGS Circular 671.
- Balthazar V, Vanacker V and Lambin E F 2012 Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas; *Int. J. Appl. Earth Obs. Geoinf.* **18** 436–450.
- Breiman L, Friedman J H, Olshen R A and Stone C J 1984 *Classification and regression trees*; Wadsworth, Belmont CA.
- Carrão H, Gonçalves P and Caetano M 2008 Contribution of multispectral and multitemporal information from MODIS images to land cover classification; *Remote Sens. Environ.* **112** 986–997.
- Census of India 2011 Government of India, New Delhi.
- Chen D and Stow D A 2002 The effect of training strategies on supervised classification at different spatial resolutions; *Photogram. Eng. Remote Sens.* **68**(11) 1155–1161.
- Dorren L K A, Maier B and Seijmonsbergen A C 2003 Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification; *Forest Ecol. Manag.* **183** 31–46.
- Duro D C, Franklin S E and Dubé M G 2012 A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery; *Remote Sens. Environ.* **118** 259–272.
- Evrendilek F and Gulbeyaz O 2011 Boosted decision tree classifications of land cover over Turkey integrating MODIS, climate and topographic data. *Int. J. Remote Sens.* **32**(12) 3461–3483.
- Feitosa R Q, Costa G A O P, Mota G L A, Pakzad K and Costa M C O 2009 Cascade multitemporal classification based on fuzzy Markov changes; *ISPRS J. Photogram. Remote Sens.* **64** 159–170.
- Friedl M A and Brodley C E 1997 Decision tree classification of land cover from remotely sensed data; *Remote Sens. Environ.* **61** 399–409.
- Friedl M A, Brodley C E and Strahler A H 1999 Maximizing land cover classification accuracies produced by decision trees at continental to global scales; *IEEE Trans. Geosci. Remote Sens.* **37**(2) 969–977.
- Friedman J H 1977 A recursive partitioning decision rule for nonparametric classification; *IEEE Trans. Comput.* **C-26** 404–408.
- Gamanya R, Maeyer P D and Dapper M D 2007 An automated satellite image classification design using object oriented segmentation algorithms: A move towards standardization; *Expert Syst. Appl.* **32** 616–624.
- Gamanya R, Maeyer P D and Dapper M D 2009 Object oriented change detection for the city of Harare, Zimbabwe; *Expert Syst. Appl.* **36** 571–588.
- Gregorio A D and Jansen L J M 2000 *Land Cover Classification System (LCCS): Classification Concepts and User Manual*; FAO, Rome, Italy.
- Hansen M C, DeFries R S, Townshend J R G and Sohlberg R 2000 Global land cover classification at 1 km spatial resolution using a classification tree approach; *Int. J. Remote Sens.* **21** 1331–1364.
- Hunt E B, Marin J and Stone P T 1996 *Experiments in Induction*; Academic Press, New York.
- Jensen J R 2005 *Introductory Digital Image Processing: A Remote Sensing Perspective*; Pearson Prentice Hall: Upper Saddle River.
- Jensen J R and Cowen D J 1999 Remote sensing of urban/suburban infrastructure and socio-economic attributes; *Photogram. Eng. Remote Sens.* **65** 611–622.
- Kandrika S and Roy P S 2008 Land use land cover classification of Orissa using multi-temporal IRS-P6 AWiFS data: A decision tree approach; *Int. J. Appl. Earth Obs. Geoinf.* **10**(2) 186–193.
- Lillesand T M, Kiefer R W and Chipman J 2007 *Remote Sensing and Image Interpretation*; 6th edn, Wiley & Sons, New York.
- Lu D and Weng Q 2007 A survey of image classification methods and techniques for improving classification performance; *Int. J. Remote Sens.* **28** 823.
- Mather P M 2004 *Computer Processing of Remotely-Sensed Images: An Introduction*; 3rd edn, Chichester, UK: John Wiley & Sons Ltd.
- Moody A and Strahler A H 1994 Characteristics of composited AVHRR data and problems in their classification; *Int. J. Remote Sens.* **15**(17) 3473–3491.
- Morgan J N and Messenger R C 1973 THAID a sequential analysis program for analysis of nominal scale dependent variables; Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor.
- Morgan J N and Sonquist J A 1963 Problems in the analysis of survey data, and a proposal; *J. Am. Stat. Assoc.* **58** 415–434.
- Murthy Sreerama K, Kasif S and Salzberg S 1994 A system for induction of oblique decision trees; *J. Artif. Intell. Res.* **2** 1–33.
- Otukey J R and Blaschke T 2010 Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms; *Int. J. Appl. Earth Obs. Geoinf.* **12** S27–S31.
- Pal M and Mather P M 2003 An assessment of the effectiveness of decision tree methods for land cover classification; *Remote Sens. Environ.* **86** 554–565.
- Pu R, Gong P, Michishta R and Sasgawa R 2008 Spectral mixture analysis for mapping abundance of urban surface components from the Terra/ASTER data; *Remote Sens. Environ.* **112** 939–954.
- Punia M, Joshi P K and Porwal M C 2011 Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data; *Expert Syst. Appl.* **38**(5) 5577–5583.
- Quinlan J R 1975 *Machine Learning*; Vol 1(1), University of Sydney.
- Quinlan J R 1993 *C4.5: Programs for machine learning*; San Mateo, California: Morgan Kaufmann Publishers.

- Shaw G A and Burke H K 2003 Spectral imaging for remote sensing; *Lincoln Lab. J.* **14**(1) 3–28.
- Stefanov W L, Ramsey M S and Christensen P R 2001 Monitoring urban land cover change: An expert system approach to land cover classification of semi-arid to arid urban centers; *Remote Sens. Environ.* **77** 173–185.
- Stehman S V 1996 Estimation of Kappa coefficient and its variance using stratified random sampling; *Photogram. Eng. Remote Sens.* **26** 401–407.
- Thapa R B and Murayama Y 2009 Urban mapping, accuracy and image classification: A comparison of multiple approaches in Tsukuba city, Japan; *Appl. Geogr.* **29** 135–144.
- Wen X, Hu G and Yang X 2008 CBERS-02 remote sensing data mining using decision tree algorithm; *Int. Conf. on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop*, Article No. 58.
- Wentz E A, Nelson D, Rahman A, Stefanov W L and Roy S S 2008 Expert system classification of urban land use/cover for Delhi, India; *Int. J. Remote Sens.* **29**(15–16) 4405–4427.
- www.suratmunicipal.gov.in

MS received 21 November 2012; revised 15 February 2013; accepted 25 February 2013