Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Scholars Research Library

Annals of Biological Research, 2012, 3 (8):4045-4053
(http://scholarsresearchlibrary.com/archive.html)

Annals of Biological Research
Scholars Research Library
ISSN 0976-1233
CODEN (USA): ABRNBW

Decision Tree Land Use/ Land Cover Change Detection of Khoram Abad City Using Landsat Imagery and Ancillary SRTM Data

Hamid Reza Matinfar[a] and Majid Shadman Roodposhti[b]

[a]*Dept. of Soil Science, Lorestan University, Lorestan, Iran*
[b]*Dept. of Geography, Tehran University, Tehran, Iran*

ABSTRACT

Change detection is a general remote sensing technique that compares imagery collected over the same area at different times and highlights features that have changed. In this paper, land cover of Khoram Abad, a city in Lorestan province of Iran, was examined in a case study via post classification technique and decision tree classifier. The Decision Tree (DT) classifier performs multistage classifications by using a series of binary decisions to place pixels into proper classes. Input data may be used from various sources and data types. Such as, multispectral data, digital elevation model (DEM) and slop to find features with similar spectral reflectance but different in elevation. In order to carry out comprehensive analysis of Khoram Abad land cover changes from years 1992 to 2009, TM data obtained from Landsat Satellite and digital elevation model of shuttle radar topography mission were used. Finally, post classification analysis using DT classifier showed notable improvement in classification accuracy in spite of high correlation of multi-spectral data.

Key word: Land Cover, Change Detection, Decision Tree, Multi-spectral Images, Khoram Abad.

INTRODUCTION

Land use refers to man's activities on earth, which are directly related to land, whereas land cover denotes the natural features and artificial constructions covering the land surface [35]. The geospatial phenomena are changing over time and the land cover information has to be up-date periodically. Up-to-date knowledge of land cover is an important tool for the various planning authorities with responsibilities for the management of territory [19]. However, it should be noted that planners and land managers require accurate data to address land cover problems. Although the priority is for land use (economic) information, land cover information is more easily mapped and can serve as an approximation of land use.

During the past decades, not only remote sensing images have become an important tool for land use classification and mapping [14] but also because of the advantages of repetitive data acquisition, they have become major data sources from local to global scales for different change detection applications [17]. There were several studies conducted to investigate land use changes during the time some of which will be referred to in this section. Tamilenthi et al [34] used principle component analysis (PCA) to examine changed pattern of urban/bare areas in 1973-2010. Sunar [25] also used five techniques, including: adding, subtracting, dividing, principle component

(PCA), and post classification analysis to detect land cover changes in Aykitali, Turkey. He found that adding and subtracting images were the most simple among these techniques while PCA and post classification analyses showed better results in change detection.

Tardi and Contalgon [33] also used three methods including: multi-temporal color composite, subtraction, and classification in order to examine physical development of Massachusett's urban area and resulting land cover changes. Finally, they used post classification analysis in order to estimate total accuracy. Qiasvand [28], also concluded similar study via PCA and subtraction techniques so as to present south Tehran land cover map and he reported that regression analysis in conjunction with PCA showed better results. Jahani [11], utilized satellite images (Spot) and normalized difference vegetation index in Tehran land cover mapping project. Consequently, on the basis of the earlier studies about land cover change detection, it is obvious that most researchers used subtraction and PCA techniques to detect changes in land cover and, in further step; by classifying multi-temporal images they showed results in quantitative form.

More recently, decision tree algorithms have been used for the classification of global datasets with promising results [9, 10, 26]. Decision tree techniques have been used successfully for a wide spectrum of classification problems in various fields [23]. They are computationally efficient and flexible, and also have an intuitive simplicity. They therefore have substantial advantages in remote sensing applications. One of the simplest alternatives to traditional classification systems is decision tree classification. The basis of this approach is establishing a set of binary rules that are applied sequentially to discriminate between different target categories. Those rules include thresholds on spectral bands, but also on auxiliary information, such as soil maps, slope, or digital elevation model, and therefore are very flexible to different types of input data.

Running et al. [30, 31, 32] and Nemani and Running [22] applied a tree-based decision structure to a global data set of NDVI values. The data set is both well understood and well behaved and the classification tree was defined solely on analyst expertise, where the threshold values are defined based on ecological knowledge. This algorithm, however, is somewhat difficult to implement since significant spatial, temporal and spectral variation make globally robust user defined threshold specification almost impossible.

More commonly, tree-based algorithms use statistical procedures, which estimate the classification rules from a training sample. A classic example is the classification and regression tree (CART) model described by Breiman et al [1]. These algorithms combine the advantages of statistically based techniques and learning algorithms, which have their origin in the machine-learning and pattern-recognition communities. Tree-based methods are supervised techniques and therefore a training set is required from which the classes can be learned.

A critical step in the estimation of a decision tree is to prune the tree back in order to avoid over fitting. By convention a tree is constructed in such a way that all (or nearly all) training samples are correctly classified, i.e. the training classification accuracy is 100%. If the training data contains errors the tree will be over fitted and will generate poor results when applied to unseen data. This study carry out on land use/land cover changes base upon remotely sensed data and decision tree technique.

2. Study Area

Khoram Abad city is located between 48° 13ʹ and 48° 23ʹ of eastern longitude and 33° 23ʹ to 33° 33ʹ of northern latitude within the center of Lorestan province, in western part of Iran. The city located in a valley and has been surrounded by Zagros Mountains. The total area of the city is 6233 Km^2 and Its Altitude from free seas is about 1134 m. Climatically, Lorestan province can be divided into three parts: the mountainous regions, such as Borujerd, Dorood, Azna, Noor Abad and Alishtar which experience cold winter and moderate summers. In the central region, the spring season begins from mid-February and lasts till about mid May. Khoram Abad city is located in this realm. The southern area is under the influence of the warm air currents of Khuzestan, have hot summers and relatively moderate winters.

Khoram Abad River which is the major river within the study area emanates from northern mountains and continues it path across the city toward west. Figure 1 shows Khoram Abad location in Lorestan province, Iran.

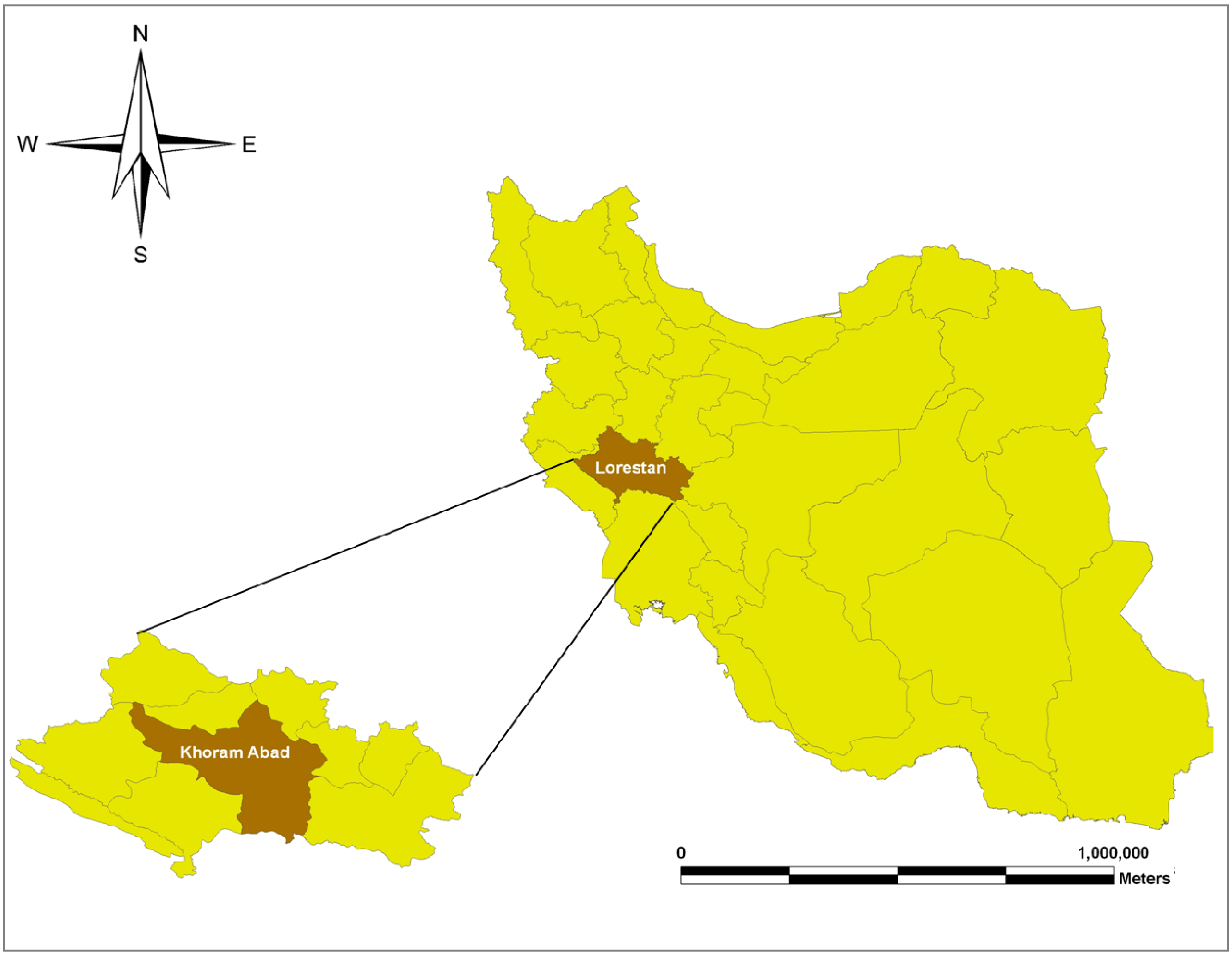

Figure 1. Map of study area

3. Data and Software

In the present study, Landsat images of Khorm Abad were acquired for two Epochs (1992 and 2009) along with digital elevation model of study area. We also used image processing software, i.e. ENVI 4.5 and geographic information system, i.e. ArcGIS 9.3 the change detection workflow. Table 1 shows data characteristics which were used in this research:

Table 1. Satellite data characteristics

Sensor Type	Imagery Date	WRS Path	WRS Row	Band No.	Radiometric Resolution	Spatial Resolution
TM	1992/8/27	037	166	1,2,3,4,5,7	7	30.00 m
TM	2009/8/7	037	166	1,2,3,4,5,7	8	28.50 m
SRTM	2002/8/7	037	166	1	1	90.00 m

MATERIALS AND METHODS

The main tools for the analysis of this study are image processing software and GIS to obtain a macro view of Khoram Abad land cover change and carry out comprehensive analysis of this change at the same time. In this study, post classification comparison upon decision tree classification, is used as a quantitative technique of analysis. The overall methodology is summarized in Figure 2.

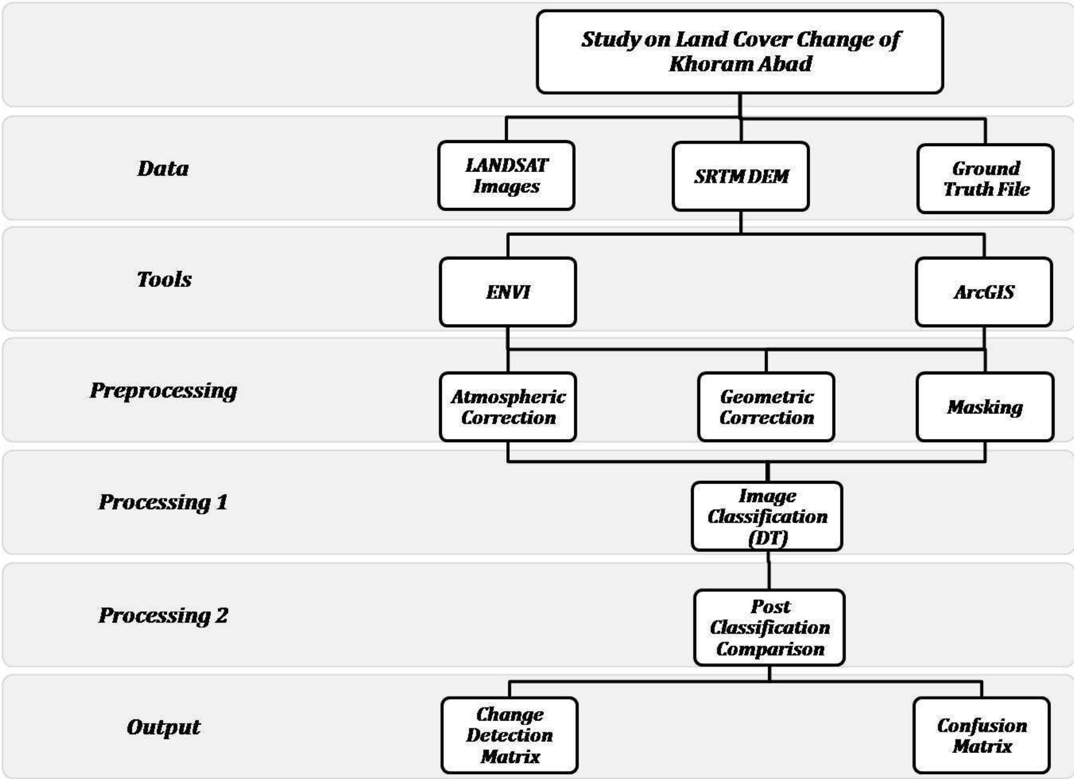

Figure 2. Diagram of methodology.

4.1 Image Preprocessing

Preprocessing of satellite images prior to image classification and change detection is essential and commonly comprises a series of sequential operations, including atmospheric correction or normalization, image registration, geometric correction, and masking (e.g. for clouds, water, irrelevant features) [5]. In the preprocessing stage, it is vital to eliminate any kind of atmospheric effects before any image analysis or information extraction are carried out [3]. This becomes especially important when scene to scene comparisons of two or several images in applications, such as change detection, are being sought [21]. Some general advice on the need for atmospheric correction in classification and change detection studies is provided by Song et al. [24]. These authors suggest that atmospheric correction is not required as long as the training data and the data to be classified are measured on the same relative scale. However, if multi-temporal image data are being processed then they must be corrected for atmospheric effects to ensure that they are comparable.

In this research, Dark Object Subtraction (DOS) is used as an approach for atmospheric correction, which is perhaps the simplest yet most widely used image-based absolute atmospheric correction approach for classification and change detection applications [12, 16]. This approach assumes the existence of dark objects (zero or small surface reflectance) throughout a Landsat TM scene and a horizontally homogeneous atmosphere. The minimum DN value in the histogram from the entire scene is thus attributed to the effect of the atmosphere and is subtracted from all the pixels [4]. It is notable that Dark object subtraction method is used only for TM image (TM image was already corrected atmospherically via USGS center and image histogram showed no offset value within the image).

In the further step, geometric registration should be done in order to prepare two or more images for comparison [18]. To conform the pixel grids and remove any geometric distortions in the TM imagery, the first TM image, August 27, 1992, was registered to TM image, UTM coordinate system Zone 39 North, based on 13 ground control points collected from the whole study area. Afterwards, first order transformation, and the nearest neighbor resampling of the uncorrected imagery was performed. First order transformation is also known as a linear transformation which applies the standard linear equation ($y = mx + b$) to the X and Y coordinates of the GCP's. The nearest neighbor resampling method uses the value of the closest pixel to assign to the output pixel value and thus transfers original data values without averaging them as other methods do; therefore, the extremes and subtleties of the data values are not lost.

Image fit was considered acceptable if the RMS error was below 15 m (RMS error <15 m) or one-half pixel wide (RMS= 0.5). Additional research obtained after the preprocessing phase was complete, indicate that due to misregistration, the accuracy of remotely sensed change detection can be substantially degraded [7]. Results of their analysis on Landsat TM data indicated that a registration accuracy of less than one-fifth of a pixel (0.2) is required to achieve a change detection error of less than 10%. However, Daie and Khorram [17] also suggest that there are inherent differences between TM image pairs which may be more or less sensitive to image misregistration than other pairs. Here, an overall RMS error of geometric correction less than 0.2 pixels was achieved for transformation (3.31 m).

Finally, in order to subset the study area from each of the two Landsat scenes, a vector file defining the city boundary with the same georeferenced coordinates as the Landsat images, UTM Zone 39 north, was imported into ENVI framework. Then the city area was masked via the city boundary vector file which was converted into a binary bitmap mask and overlaid on to each of the TM scenes.

4.2 Image Classification Using Decision Tree Classifier

As opposed to single stage classifiers in that only one decision is made about a pixel, as a result of which it is labeled as belonging to one of the available classes or is left unclassified. Multistage classification techniques are also possible in which a series of decisions is taken in order to determine the correct label for a pixel. The more common multistage classifiers are called decision trees. They consist of a number of connected classifiers (or decision nodes) none of which is expected to perform the complete segmentation of the image data set. Instead, each component classifier only performs part of the task [29] (Figure 3).

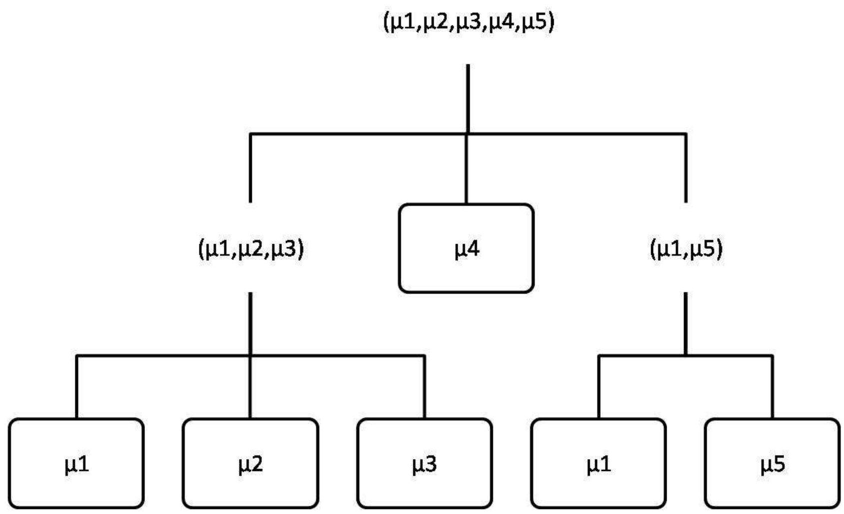

Figure 3. General decision tree hierarchy

The advantages of using a multistage or tree approach to classification include that different data sources i.e. spectral data or GIS data e.g. elevation data, etc. different sets of features, and even different algorithms can be used at each decision stage. Minimising the number of features to use in a decision is significant for reducing processing time and for improving the accuracy of small class training [29]. Figure 4 shows decision tree input data including: spectral and elevation data which were used in present research to improve the classification accuracy and learning algorithm.

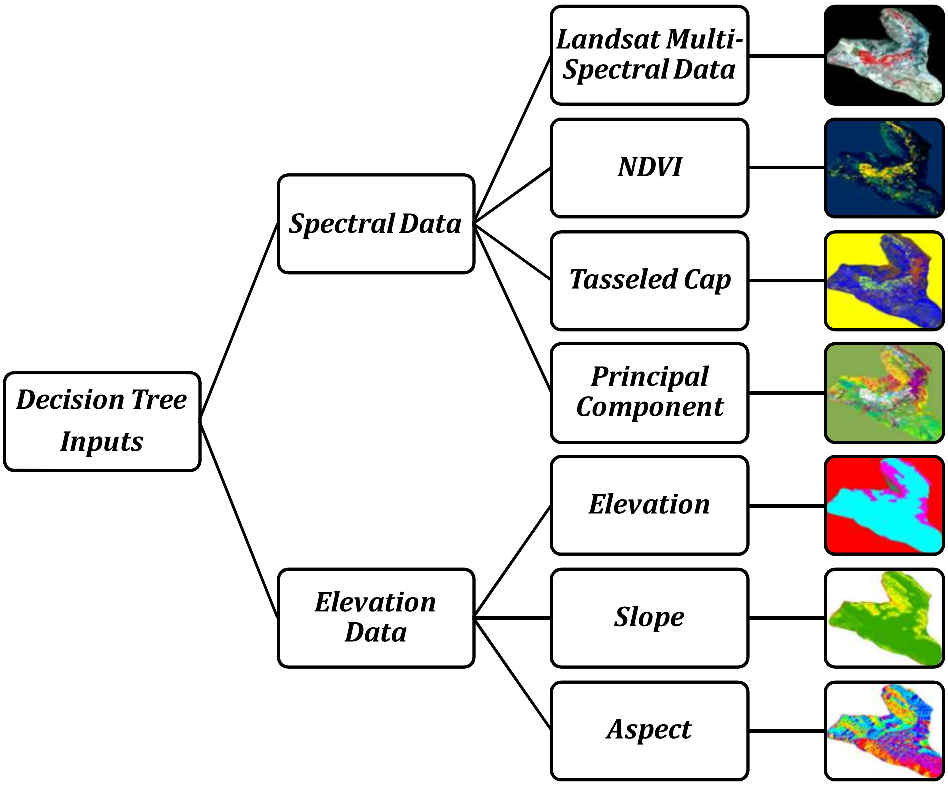

Figure 4. Decision tree inputs including spectral and elevation data

4.3 Post Classification Comparison

This is the most straightforward method of change detection. It involves the overlay (or "stacking") of two or more classified images. Change areas are simply those areas which are not classified the same at different times.

The post classification comparison method is one of the most widely used methods of remote sensing change detection. Some of the main advantages of this method are as follows: there is no need for radiometric co-registration of images involved in the analysis [13]; its sensitivity to the spectral variations due to the difference in the soil moisture, vegetation, and phonology is lower than that of the spectral change detection methods [20]; its provision of "from-to" change information [13]; it's very high change detection accuracy [20]. Two main disadvantages of this method are its dependency on the accuracy of individual classification results and also its being quite time consuming due to the classification processes it uses for all the data.

The classification scheme includes the following classes: bare Rocks, high forests, healthy vegetation, built up lands, roads network, water, farm A, farm B, bare lands. The overall accuracy of classification scheme for each of two images was estimated by means of the standard accuracy assessment procedure (i.e., an error matrix) and the following formula:

$$OA = \frac{1}{n}\sum_{k=0}^{n} p_{ii}$$

Where *OA* is overall accuracy, *n* is the total number of ground truth pixels and p_{ii} is the diagonal element of the confusion matrix.

Because of inherent deficiency of overall accuracy criterion which is correctly assigned pixels may have been assigned by chance and not based on the classification decision rule, some authors prefer to use the kappa coefficient as a measure of map accuracy [15, 6]. The kappa coefficient *k* is another measure of the accuracy of the classification and is defined by:

$$k = \frac{n \sum_{i=1}^{r} p_{ii} - \sum_{i=1}^{r} (p_{i+} * p_{+i})}{n^2 - \sum_{i=1}^{r} (p_{i+} * p_{+i})}$$

Where k is kappa coefficient, n is the total number of pixels in all the ground truth classes, r is the number of rows in the confusion (error) matrix, p_{ii} is the sum of the confusion matrix diagonals, p_{i+} is the total observations in row i and p_{+i} is the total of observations in column i.

Finally, after image classification process the classified images were overlaid by means of the same coordination and projection systems and the accurate percent of the change over this period of time (1992-2009) was calculated for each class by subtraction technique.

RESULTS AND DISCUSSION

Here, in this study attempted to identify urban Land use/ Land cover change of Khoram Abad within 1991 and 2009. Remote sensing has the capability of monitoring such changes, extracting the change in classified information from satellite data. The image classification for TM, 1992 has been done with approximate over all accuracy of 81.2%, while over all accuracy of 84.4% was obtained for TM, 2009 (Figure 5 and 6).

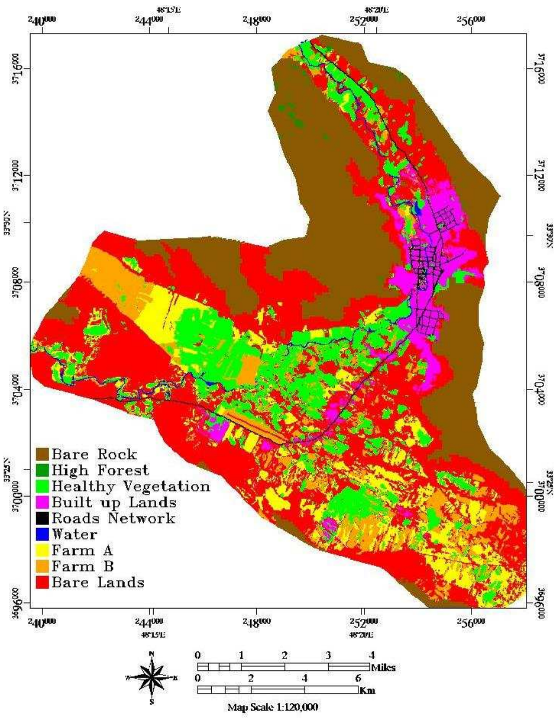

Figure 5. The classified map of TM image (August 27th, 1992) with Overall Accuracy = (203/250) 81.2% and Kappa Coefficient = 0.802

It should be noted that the accuracy of the post-classification comparison is totally dependent on the accuracy of the initial classifications. The final accuracy very closely resembles that resulting from the multiplication of the accuracies of each individual classification [27].

This case study presents encouraging results about the change detection process. Even though, multi-spectral data are highly correlated and some land cover types have very similar spectral characteristics, i.e. some classes have a constant low reflectance over the whole spectral range with no or only minor distinct absorption features. For example, problems of misclassification occur between buildings and roads which are caused by spectral similarities between materials covering these surfaces and the influence of shadow [2, 8] or inherent problems of data sets with spatial resolution of 30 m which limits the outcome of any spectral analysis (Small road networks could not be detected precisely in Landsat data due to the mixed spectral signal).

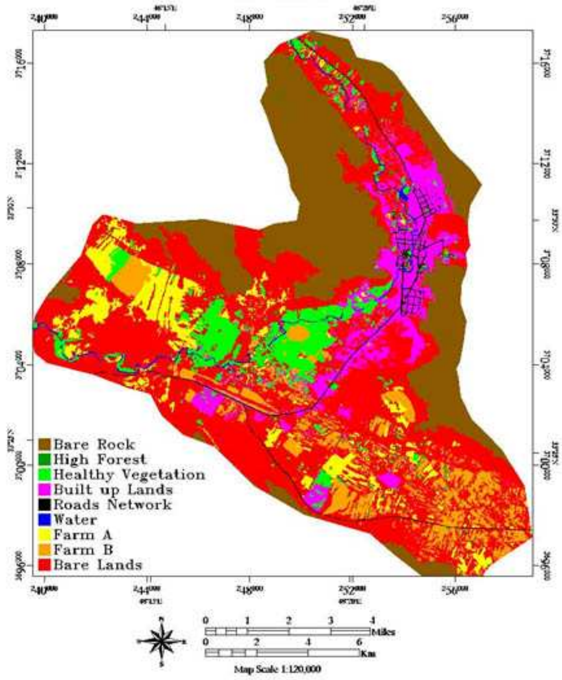

Figure 6. The classified map of TM image (August 7th, 2009) with Overall Accuracy = (2116/250) 84.4% and Kappa Coefficient = 0. 841

However, by using a knowledge based classification algorithm (DT) and some ancillary data (i.e., SRTM DEM) the classification accuracy is highly increased in all classes. The total result of this change detection analysis during mentioned time period is also shown in table 4.

Table 2. Change Detection Statistics

Class Name	Bare Rock	High Forest	Healthy Vegetation	Built up Lands	Roads Net	Water	Farm A	Farm B	Bare Lands
Bare Rock	95.84	90.82	0.11	0.08	0.00	0.00	0.05	0.52	1.41
High Forest	0.17	6.72	0.00	0.00	0.00	0.00	0.03	0.00	0.00
Healthy Vegetation	0.01	0.328	33.34	1.26	3.36	22.27	4.51	9.83	2.49
Built up Lands	0.44	0.65	8.12	76.42	39.22	12.12	3.32	3.68	6.06
Roads Net	0.00	0.00	0.58	6.77	40.38	1.25	0.76	0.50	0.70
Water	0.000	0.00	1.86	0.56	1.08	34.65	0.30	0.45	0.33
Farm A	0.002	0.00	12.06	0.39	0.35	1.30	24.21	11.32	5.43
Farm B	0.03	0.00	8.59	1.33	2.51	1.46	19.21	31.82	5.52
Bare Lands	2.35	0.16	35.30	13.13	13.07	26.60	47.28	41.84	77.95
Class Change	4.16	93.27	66.66	23.57	59.61	65.34	75.79	68.17	22.045
Image Difference	0.07	-72.29	-48.33	62.46	22.90	0.10	-28.05	-5.45	17.23

In the above table the "Class Changes" row indicates the total number of initial state pixels that changed classes. For example, the total class changes for high forest and bare rocks are 93.279 and 4.162 percent respectively. In other words, 93.279 percent of high forest class is lost (mainly changed to bare rocks class) which is the highest change rate among all classes and only 4.162 percent of bare rocks class is changed which is the least. As can be seen from Table 2, the obvious changes in the study area, mostly owed to the human activities and interventions in agricultural and natural environments which are displaying in many ways.

Finally the "Image Difference" row is the difference in the total number of equivalently classed pixels in the two images, computed by subtracting the initial state class totals from the final state class totals divided by initial state class totals. An image difference that is positive (i.e. in bare rocks, built up lands, road networks, water and bare lands) indicates that the class size increased and if image difference that is negative (i.e. high forest, farm A and farm B) it indicates that the class size decreased. Here, image difference of change detection analysis shows that built-up lands has been the most increasing class by 62.46 percent, mainly because changing demands of increasing population and high forest has been the most decreasing class decreased by -72.29 percent.

REFERENCES

[1] Breiman L, Friedman J, Olshen R, and Stone C, Classification and Regression Trees. Wadsworth International Group, Belmont, Canada, **1984**.
[2] Carleer AP, Wolff E, *International Journal of Remote Sensing*, **2006**, 27(5-6), 1035-1051.
[3] Chavez PS, *Remote Sensing of Environment*, **1988**, Vol. 24, no.3, pp.459-479.
[4] Chavez PS, *Remote Sens*. **1989**, 55:1285–1294.
[5] Coppin PR and Bauer ME, *Remote Sensing Reviews*, **1996**, 13: 207-234.
[6] Congalton RG and Green K, Assessing the Accuracy of Remote Sensing Data: Principles and Practices, Lewis Publications, Roca Baton, Florida, **1999**, 137 p.
[7] Dai X L and Khorram S, *IEEE Transactions on Geosciences and Remote Sensing*, **1998**, v36 i5. 1566-1577.
[8] Dare PM, *Photogrammetric engineering and remote sensing*, **2005**, 71(2), 169-177.
[9] DeFries RS, Hansen M, JRG Townshend and R Sohlberg, Global Land Cover Facility, University of Maryland Institute for Advanced Computer Studies, College Park, Maryland, U.S.A. **1998**.
[10] Friedl M, and Brodley C, *Remote Sensing of Environment* , **1997**,61, 399-409.
[11] Jahani A, *International Journal of Remote Sensing*, **1997**, 27(5-6), 1035-1051.
[12] Jakubauskas ME, *Remote Sensing of Environment*, **1996**, 56:118–132.
[13] Jensen JR, A remote sensing perspective, 3rd ed., Ch. 12: Digital Change Detection, **2005**, p. 467 – 494.
[14] Homer CG, Ramsey RD, Edwards TC, and Falconer A, *Photogrammetric Engineering and Remote Sensing*, **1997**, 63, 59–67.
[15] Hudson WD, and Ramm CW, *Photogrammetric Engineering and Remote Sensing*, **1987**, 53, 421–422.
[16] Huguenin RL, Karaska MA, Blaricom DV, and Jensen JR, *Photogramm. Eng. Remote Sens*. **1997**, 63(6): 717–725.
[17] Lu D, Mausel P, Brondizio E and Moran E, *International journal of remote sensing*, **2003**, 25(12), 2365-2407.
[18] Macleod RS and Congalton RG, *Photogrammetric and Engineering Remote Sensing*, **1998**,64(3): 207-216.
[19] Marçal ARS, Borges JS, Gomes JA and Pinto Da Costa JF, *International journal of remote sensing*, **2005**, 26(7), 1347-1362.
[20] Mas JF, *International Journal of Remote Sensing*, **1999**, Vol. 20, # 1, 139 – 152.
[21] Mather PM , Computer Processing of Remotely-Sensed Images: An Introduction, Translated by M. N, SAMT Publications, **1997**, p. 100 – 150.
[22] Nemani RR and Running SW, *Ecological Applications*, **1997**, 7(1), 79−90.
[23] Safavian SR and Landgrebe D, *IEEE Transactions on Systems, Man, and Cybernetics* 21, **1991**, 3, 660674
[24] Song C , Woodcock CE , Seto KC, Pax Lenney M and Macomber SA, *Remote Sens. Environ*. **2001**, 75:230-244.
[25] Sunar F, *International Journal of Remote sensing*, **1998**, 19:245-265.
[26] Swain PH and Hauska H, *IEEE Trans. Geosci. Remote Sensing*, **1977**, vol. GE-15, pp. 142–147.
[27] Petit CC and Lambin EF, *International journal of geographical information science*, **2001**, 15(8), 785-803.
[28] Qiasvand Q, MS. C. thesis, Tarbiat Modares University Tehran, Tehran, Iran, **1997**, pp. 56-62
[29] Richards JA and Jia X, Remote sensing digital image analysis: an introduction, 4th ed., Springer, Berlin, **2005**, p. 321 – 323.
[30] Running S and Hunt E, Scaling Physiological Processes: Leaf to Globe, Ehleringer, J. and Field, C., **1993**.
[31] Running S, Justice C, Salomonson V, Hall D, Barker J, Kaufmann Y, Strahler A, Huete A, Vanderbilt JMV, Wan Z, Teillet P, and Carneggie D, *International Journal of Remote Sensing*, **1994**, 15, 3587-3620.
[32] Running SW, Loveland TR, Pierce LL, Nemani RR, and Hunt ER, *Remote Sensing of Environment*, **1995**, 51, 39-48.
[33] Tardi PS and Congalton RG, Department of Natural Resources. University of New Hampshire Durham, **2001**, pp. 156-162
[34] Tamilenthi S, Punithavathi J, Baskaran R and ChandraMohan K, *Archives of Applied Science Research*, **2011**, 3 (1):277-286
[35] Singh P and Singh S, *Archives of Applied Science Research*, **2011**, 3 (5):10-16.