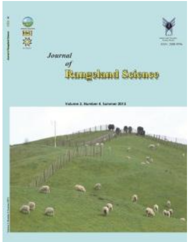
Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Contents available at ISC and SID

Journal homepage: www.rangeland.ir



Full Length Article:

Application of Different Methods of Decision Tree Algorithm for Mapping Rangeland Using Satellite Imagery (Case Study: Doviraj Catchment in Ilam Province)

Marzban Faramarzi[A], Hassan Fathizad[B], Nasibe Pakbaz[C], Behzad Golmohamadi[D]

[A]Assistant Professor, Rangeland and Watershed Management Group, Faculty of Agriculture, Ilam University, Ilam, Iran (Corresponding Author), Email: faramarzi.marzban@gmail.com
[B]M.Sc. Graduate of Combating Desertification, Faculty of Agriculture, Ilam University, Ilam, Iran
[C]M.Sc. Graduate of Agronomy, Agriculture College, Ilam University, Ilam, Iran
[D]M.Sc. Graduate of Rangeland Management, Faculty of Natural Resources, Tarbiat Modares University, Iran

Received on: 06/10/2013
Accepted on: 06/01/2014

Abstract. Using satellite imagery for the study of Earth's resources is attended by many researchers. In fact, the various phenomena have different spectral response in electromagnetic radiation. One major application of satellite data is the classification of land cover. In recent years, a number of classification algorithms have been developed for classification of remote sensing data. One of the most notable is the decision tree. The aim of this study was to compare three types of decision trees split algorithm for land cover classification in Doviraj catchment in Ilam province, Iran. For this, propose, first, the geometric and radiometric corrections were performed on the 2007 ETM+ data. Field data as training sites were collected in the various classes of land use. The results of image classification accuracy assessment showed that the Gini split classification. With kappa value 89.98 and the entire accuracy 91.17% was significantly higher, then categorization of branching and the branching ratio and Entropy with kappa values of 88.45 and 90.65 and the entire accuracy of 86.21 and 86.15%, respectively.

Key words: Classification tree, Gini, Entropy, Ratio, Doviraj

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Introduction

Some of the phenomena and ground effects such as rangeland are changed because of natural or human activities over time. This condition influences ecosystem function and conditions. Therefore, the need for detection, prediction, and observation of such changes in an ecosystem is of great importance. In addition, gaining knowledge about the rangeland condition and its health plays an important role in management factor. Evaluation and monitoring of the rangeland condition is often dificult by only field data collection in global and regional scale. In this method, field data are limited to small places and short time intervals (Pettorelli *et al.*, 2005). Remotely sensed evaluation is a very useful technology that can be used to obtain information from a distance without coming into direct contact with the object of interest, e.g. rangeland. As well, satellite imagery data is rapidly and repeatedly over large areas with high accuracy information. Accordingly, many researchers have employed this data to investigate plant (Mokhtari *et al*., 2000; Huete, 2004).

Since the main objective of the satellite images processing is providing efficient and thematic maps, the selection of convenient method of classification plays a great role in so doing. Currently, there are different types of classification methods. Conventional methods of classification use statistical techniques including classification methods of maximum likelihood, minimum distance, and any expression that may serve which employed parametric classification algorithm. Statistical classification methods are depended upon data model such as normal distribution and thus the efficiency of these methods depend on the agreement amount of the data with this model. If the input data distribution is almost normal, the efficiency of statistical classification methods can be good. Despite the limitations of this method which is from the normal distribution assumption of the class signature (Swain and Davis, 1978), this is perhaps one of the mostly used classification methods (Wang, 1990; Hansen *et al*., 1996). Decision tree methods, unlike other classification approaches (for example: maximum likelihood method or artificial neural network methods) that simultaneously use a series of features (bands) for classification process in a single stage, are based on a series or multi-stage decision plan (Xu *et al*., 2005).

This method has been used successfully for a wide range of problems including image classification of remote evaluation (Yang *et al*., 2003; Xu *et al*., 2005; Chubey *et al*., 2006). There are numerous articles regarding the classification of vegetation coverage and lands use using remote sensing evaluating data. The first land Earth coverage classification at global-scale with the maximum likelihood method has been done by Defris and Townshend (1994) and the provision of land coverage map with resolution power of 1 Km using unsupervised classification approach has been conducted by Loveland *et al.* (2000). In recent years, because of the limitations of these methods, decision tree and neural network approaches having nonlinear and nonparametric characteristics have been used in regional and global levels. Hansen *et al*. (1996) have employed the NOAA/AVHR data to provide land Earth coverage map at global scale and with spatial resolution power of 10 x 10 as well as decision tree and maximum likelihood methods. The accuracy of the decision tree method has been reported better than maximum likelihood (Hansen *et al.*, 1996). The present study made an attempt to compare various methods of decision tree to extract the map of rangeland using Ladnsat ETM^+ satellite imagery as well as RS and GIS technology.

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Materials and Methods

Study area

The study area is located in the southern part of the Doviraj catchment in Ilam province, Iran with an area of 31938 ha (32°34´50"-32°46´54" N; 47°23´48"-47°39´11" E) (Fig. 1). The annual rainfall, evaporation, and temperature are 264.4 mm, 3117 mm, and 31.4°C, respectively (Shahriari *et al.*, 2010). (Fig. 2), demonstrates the process of the research stages.

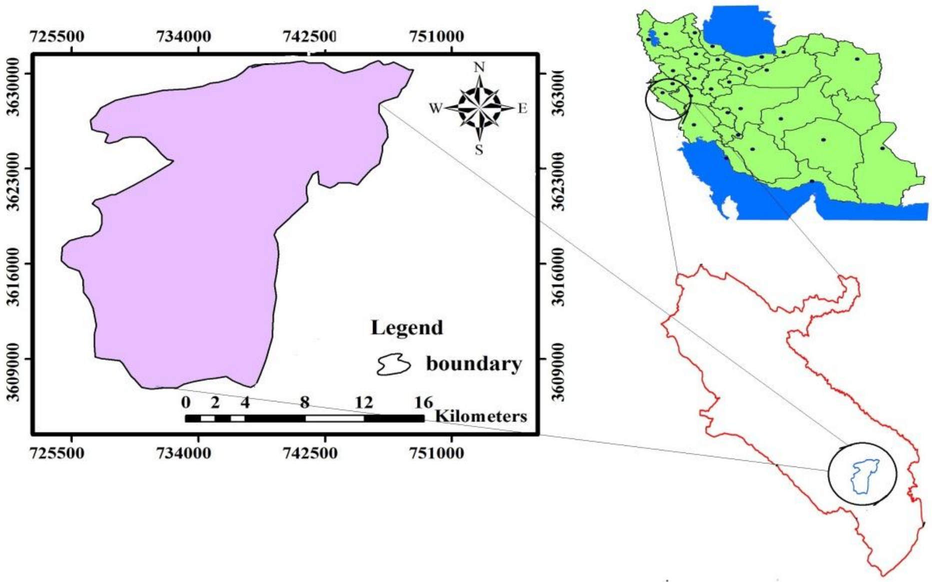


Fig. 1. Map of the study area (Doviraj catchment) in Ilam province

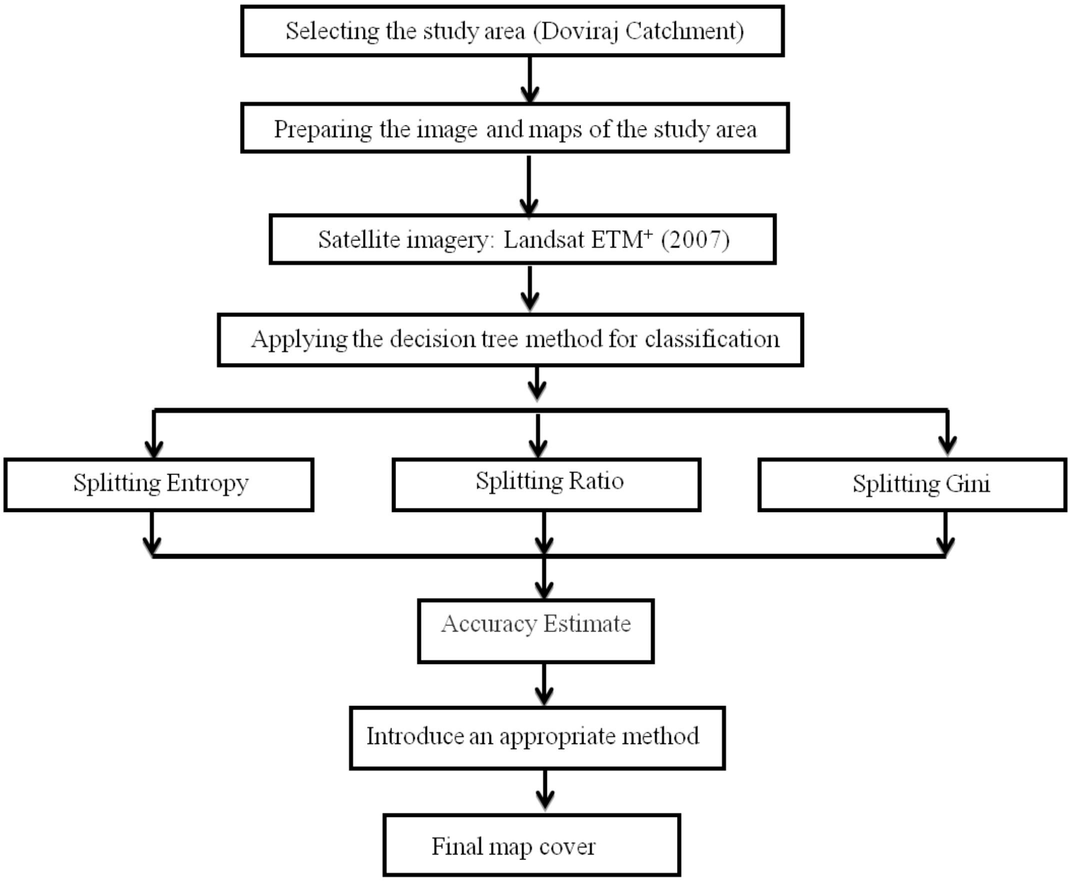


Fig. 2. Flowchart of the research stages

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Decision tree

Decision tree goes through data sequential separation in every node to new nodes containing more homogeneous subsets based on the training pixel. The newly formed node may create a leaf in a case that the training pixels contain only one class or most of the pixels with one class. When there is no other node to split (separation), the final rules of tree decision (classification) would be formed (Fig. 3). Idrisi 15 software was used for this analysis.

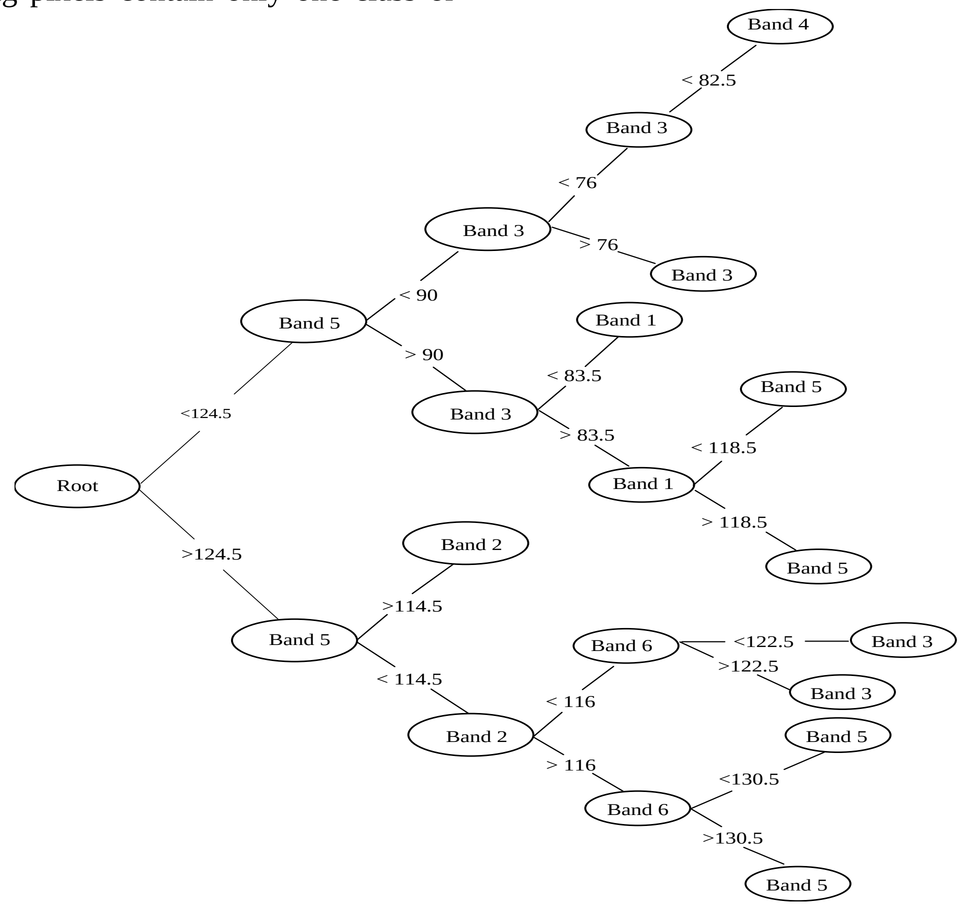


Fig. 3. A graphical design of decision tree

Entropy splitting method

it is based upon the following equation (Equation 1):

$$Entropy = -\sum_{j=1}^{k} \frac{freq(C_J, S)}{|S|} \times \log_2\left(\frac{freq\,(C_{j,}S)}{|S|}\right)$$

(Equation 1)

Where

$|S|$: number of pixels in the group S.

Cj: number of class j pixels in the group S.

k: number of groups from j=1 to k

Gain ratio splitting method

The gain ratio algorithm tries to overcome to the potential distortion (earsplitting) of entropy algorithm via the process of normalization. If we define the (X) splitting information as follows (Zambon *et al.*, 2006), the split info represents the potential information generated by dividing S group to the n subgroup (Equation 2):

$$\text{Splitinfo}(X) = -\sum_{i=1}^{n} \frac{|S_i|}{S} \times \log_2\left(\frac{|S_i|}{|S|}\right)$$

(Equation 2)

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Where

$|S|$: number of pixels in the group S.

n: number of subgroup from i = 1 to n.

$|Si|$: number of pixels in the group S_i.

Gini splitting method

Gini splitting method tries to find the most homogeneous set among the data series and to separate it from the rest of the data (Zambon *et al.*, 2006) (Equation 3):

$$Gini(S)=\sum_{i} freq(C_{j,}S)\times(1-freq(C_j,S))$$

(Equation 3)

Where

Cj:number of class j pixels in the group S

S: group of pixels

J: group number

I: number of pixel

Geometric correction

The raw and primary images of satellite data would be had some error due to various reasons such as earth circulation and change in geometry satellite elevation and in this case, the satellite data are not comparable to each other. Therefore, the aim of geometric correction is compensating the aforementioned distortions that causes the geometric image be closer to the real world as much as possible (Alavipanah and Valdani, 2010). For this, the satellite image of Landsat ETM^+ (2007) was transformed in UTM system (Zone 38, WGS 84) using ENVI 4.5 software. Coordinates of ground control points were obtained by the following methods those of the topography map (1:50000), aerial photography (1:20000), Google Earth satellite images, and ground reference points using GPS (Fig. 4). The image with the proper distribution of control points and the Root Mean Square Error (RMSE) of approximately 0.3 pixels were georeferenced. Finally, re-sampling method of the nearest neighbor was used to determine the new values of pixels and then the following classification methods for Landsat ETM^+ satellite image classification was employed. Three methods of decision tree were used for this study, namely Entropic Splitting method, Gini Splitting method, ande Gain Ratio Splitting method.

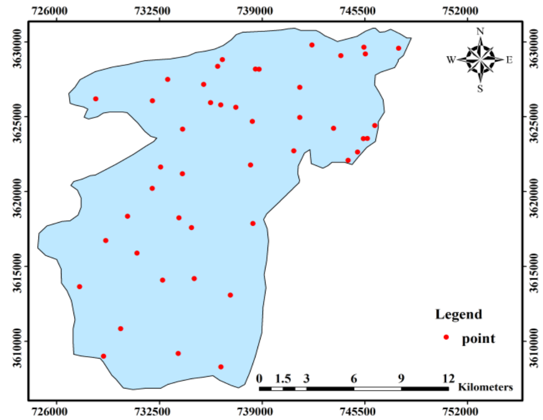


Fig. 4. Location of georeferenced points for geometric correction

Classification accuracy valuation

The accuracy estimation is important to understand the obtained results and to apply the results to make decisions. Overall accuracy, producer's accuracy, user's accuracy, and Kappa coefficient are the most common parameters of accuracy estimation (Lu *et al.*, 2004; Alavipanah, 2005; Bonyad and Hajighaderi, 2007). Theoretically, overall accuracy probabilities cannot be a good scale for evaluating classification results because the role of chance is of significant in this index. The overall accuracy is calculated through the sum of the main diagonal elements of the error matrix divided by the total number of pixels according to the (Equation 4) (Alavipanah, 2005).

$$OA=\frac{1}{N}\sum P_{ii} \qquad \text{(Equation 4)}$$

Where

OA: Overall Accuracy

N: Number of experimental pixels

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Σp_{ii}: Sum of the main diagonal elements of the error matrix

Due to the discrepancies on the overall accuracy, Kappa index is often used in administrative works when the comparison of classification accuracy is taken into account since Kappa index considers the incorrectly classified pixels. Kappa index is calculated from the (Equation 5) (Bonyad and Hajighaderi, 2007).

$$Kappa = \frac{p_o - p_c}{1 - p_c} \times 100 \quad \text{(Equation 5)}$$

Where

Po: Properly observed

Pc: Prospect contract

In this stage, the ground real map with field survey was performed using stratified random sampling. After the conformity of the produced map with the ground real map, the locations (points) were determined randomly on the map. Then through desert actions, coordinates of all locations were recorded by GPS. The table of the error matrix was formed and quantitative accuracy and Kappa coefficient, which expresses the user's accuracy and producer's accuracy, were studied.

Results

To classify ETM$^+$ satellite image, lands use classes in 5 groups of forest lands class, barren lands, poor rangeland, fair rangeland, and good rangeland were determined. Then, training samples from the area level were collected using Google Earth satellite images and field visit (Fig. 5 and Table 1).

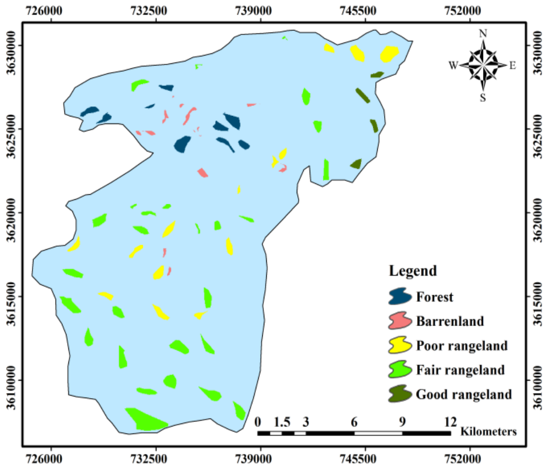


Fig. 5. Training sites of the study area

Table 1. Area of training sites of

Land Cover	Area of Training Site (ha)	Percentage of Training Site
Forest	288.3	0.9
Barren land	107.7	0.33
Poor rangeland	422.4	1.32
Fair rangeland	1008.9	3.16
Good rangeland	99.9	0.31
Total	1927.2	6.02

After identifying the separation amount of the classes, the classification of three methods of decision tree, namely, gain ratio, Gini, and entropy were measured. Therefore, the lands coverage maps for 2007 were obtained (Figs. 6, 7 and 8). In the next stage, through field operations, statistical parameters of producer's accuracy, user's accuracy, overall accuracy, and Kappa coefficient were extracted using 1:20000 aerial photographs, Google Earth satellite images, and random sampling of the under study area level (Tables 2 and 3).

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

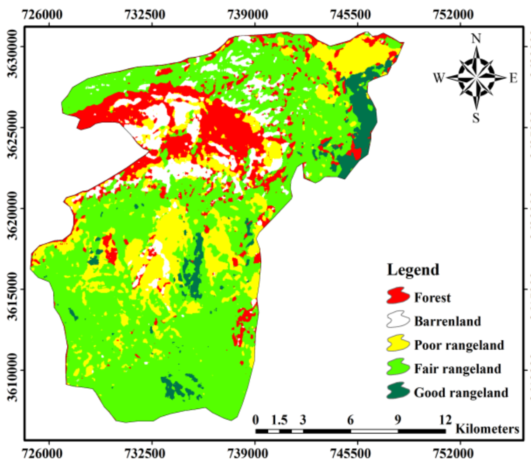


Fig. 6. Classification using Gini splitting method

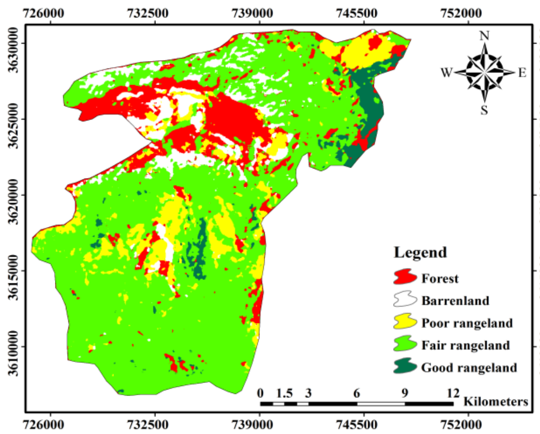


Fig. 7. Classification using Gain ratio splitting method

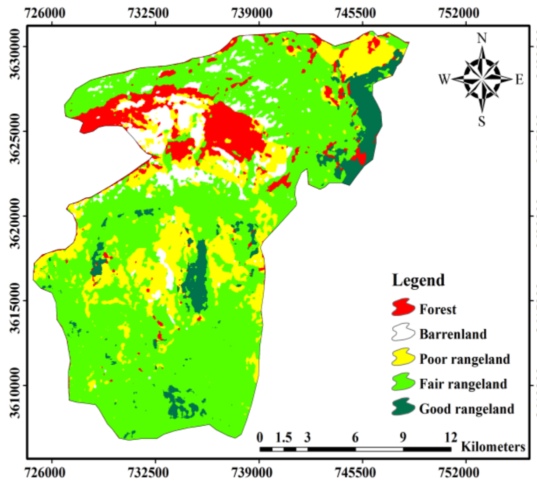


Fig. 8. Classification using entropy splitting method

When the results of (Tables 1 and 2) are analyzed, several important conclusions can be drawn: First, it was observed that forest and barren lands classes were classified with producer's accuracy higher than 90% in all the three methods. This shows the capability of high spectral resolution for these classes. Second, according to the observed results, the lowest producer's accuracy belonged to good rangeland class. This class is classified with 59.74% of producer's accuracy for the image of this area (in the case of using entropy method). While the Gini classification method (84.79%) and the Gain ratio classification splitting methods (88.11%) indicated a better producer's accuracy for this class.

Other results showed that the user's accuracy for the forest and barren lands classes was higher than 90% in three methods. The lowest user's accuracy was for good rangeland. This class was classified with the accuracy of 73.21% having a lower accuracy in compersion with other classes. The reason of this issue can be complexity or proximity of the boundaries resulted from the high spectral similarity with other classes and the mixed pixels in the experimental and training samples.

In this research, as demonstrated in the (Tables 3), the user's and producer's accuracy had a pretty good accuracy for the classification of good, fair, and poor rangeland classes. Producer's accuracy for the three pasture coverage was above 80% (except for the good rangeland in entropy method) that was a proper number for classification accuracy. Moreover, user's accuracy for the classification of good, medium, and poor rangeland class was upper than 80% and it reached above 99% for the good rangeland class in entropy method. This indicates the ability of high spectral resolution for these classes. The lowest user's accuracy was for poor rangeland class. This class was classified with 72.21% producer's accuracy for the image of this area. The reason seems to be have a similar spectral behavior or something very close to soil.

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

The classification results using three methods of the Gini, Gain ratio, and entropy have attained overall accuracy of 91.17, 90.65, and 86.21 as well as Kappa coefficient of 89.98, 88.45, and 86.15 percent, respectively. Finally, it can be said that among the three splitting methods used in this study, the Gini splitting method had a better performance.

Tabel 2. Statistical characteristics of producer's accuracy for ETM$^+$ image classification using three methods of decision tree

Land Use	Statistics of Producer's Accuracy			Statistics of User's Accuracy		
	Ratio	Entropy	Gini	Ratio	Entropy	Gini
Forest	91.33	96.74	96.97	97.74	96.84	97.43
Barren land	98.73	100	99.82	96.82	94.44	97.17
Poor rangeland	93.13	90.37	86.50	72.21	78.66	85.26
Fair rangeland	80.97	83.90	87.80	93.76	81.83	88.16
Good rangeland	88.11	59.74	84.79	90.20	99.80	88.67

Tabel 3. Accuracy of different classification method for the extracted maps from ETM$^+$ images

Algorithm	Overall Accuracy	Kappa Coefficient
Gini	91.17	89.98
Ratio	90.65	88.45
Entropy	86.21	86.16

Discussion

For the classification of ETM$^+$ satellite image, the vegetation classes were determined in five groups those of forest, barren lands, poor, fair, and good rangelands. Then, the classification was occurred based on the three methods of decision tree, namely, gain ratio, Gini, and entropy. In this study, the results of the accuracy evaluation of the classified images showed that the Gini splitting classification method with the Kappa coefficient of 89.98 and the overall accuracy of 91.17% had the highest accuracy. After that, Gain ratio and entropy splitting classifications with Kappa coefficient of 88.45 and 90.65 and the overall accuracy of 86.21% and 86.15%, respectively, were located. The results of this study are in the line with the results of Yang *et al*. (2003), Xu *et al*. (2005), and Otukei and Blaschke (2010) and in contrast with the results of Borak and Strahler (1999). Additionally, the results showed that the rangeland classification by satellite images had a high and acceptable accuracy. The user's and producer's accuracy of poor, fair, and good rangeland in most of the methods was higher than 80%.

In this study, the highest accuracy was related to Gini splitting method. This result is against what Breiman *et al*. (1984) obtained as they stated that there was no significant difference among the different splitting methods.

This study revealed that the decision tree method had many advantages to other classification methods like fuzzy Artmap artificial neural network and the maximum likelihood. As well, they were fast in terms of computation (unlike the artificial neural network methods) and they did not follow the statistical assumptions regarding the data distribution (unlike the maximum likelihood method). Arekhi (2012) showed that decision tree was accurate method for mapping land use. At last, our result demonstraded that the decision tree seemed to be proper alternative to other classification methods for rangeland.

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

Literature Cited

Alavipanah, S. K., 2005. Application of remote sensing evaluating in earth sciences. Tehran University Press, p. 478. (In Persian).

Alavipanah, S. K., and Valdani, M., 2010. Remote sensing evaluating and GIS, Tehran University Press, pp. 149-153. (In Persian).

Arekhi, S., 2012. assesment of decision tree method for mapping land use using sattelite imagery in the Chamgardelan in Ilam Province, *Jour. Geography and Land Use Planning*, 2 (4):75-84.

Bonyad, A. I., and Haji Ghaderi, T., 2007. The provision of Zanjan province natural forest map using Landsat 7 ETM+ sensor data. *Jour. Agriculture Technologies and Sciences and Natural Resources*, 42(11): 627-638. (In Persian).

Borak, J. S., and Strahler, A. H., 1999. Feature selection and land cover classification of a MODIS-like dataset for a semiarid environment. *INT. Jour. Remote Sensing*, 20: 919-938.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984. Classification and regression trees. Monterey, *CA: Wadsworth*, 358 p.

Chubey, M. S., Franklin, S. E., and Wulder, M. A., 2006. Object-based Analysis of Ikonos-2 Imagery for Extraction of Forest Inventory Parameters. Photogra*mmetric Engineering & Remote Sensing*, 72 (4): 383-394.

Defries, R. S., and Townshend, J. R. G., 1994. Global land cover: comparison of groundbaseddata sets to classi. cations with AVHRR data. In Environmental Remote Sensingfrom Regional to Global Scales, edited by G. M. Foody and P. J. Curran (New York: Wiley. 84–110.

Hansen, MC., Dub ayah, R. and Defries, R.S., 1996. Classification trees: an alternative to traditional land cover classifers. *Int. Jour. Remote Sensing*, 17:1075-1081.

Huete, A., 2004. Remote Sensing for Natural Resources Management and Enviromental Monitoring: *Manual of Remote Sensing 3 ed,* Vol.4. Univercity of Arizona.

Loveland, T. R., Reed, B.C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W., 2000. Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data. *International Jour. Remote Sensing*, 21: 1303-1330.

Lu, D., Mausel, P., Brondi´zio, E., and Moran, E., 2004. Change detection techniques. *Int. Jour. Remote Sensing*, 25(12): 2365–2407.

Mokhtari, A., Feiznia, S., Ahmadi, H., Khawajaldin, S. J., and Rahnema, F. A., 2000. Application of remote sensing evaluating in data layers provision of lands use and land coverage in soil erosion model. MPSIAC, *Jour. Construction and Research*, 46: 82-87. (In Persian).

Otukei, J. R., and Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Jour. Applied Earth Observation and Geoinformation*, 12: 27S31.

Pettorelli. N., Vik, J. O., Mysterud. A., Gaillard. J. M., Tucker, C. J., and Stenseth, N. C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Jour. Trends in ecology and evolution*. 20(9): 503–510.

Shahriari, A., Gholami, H., Fakhireh, A., Arkhi, S., and Nouri, S., 2010. Comparison of different methods for monitoring vegetation coverage in desert area of Dehloran Ein Khosh using RS and GIS. Geomatics National Conference. Iran Mapping Organization. (In Persian).

Swain, P. H., and Davis, S. M., 1978. Remote Sensing: the Quantitative Approach (New York: McGraw-Hill).

Wang, F., 1990. Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 28: 194-201.

Xu, M., Watanachaturaporn, P., Varshney, P. K. and Arora, M. K., 2005. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ*., 97: 322-336.

Yang, C., Prasher, S. O., Enright, P., Madramootoo, C., Burgess, M., Goel, P. K., and Callum, I., 2003. Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, 76: 1101–1117.

Zambon, M., Lawrence, R., Bunn, A., and Powell, S., 2006. Effect of alternative splitting rules on image processing using classification tree analysis. *Photogrammetric Engineering and Remote Sensing*, 72(1): 25–30.

Simpo PDF Merge and Split Unregistered Version - http://www.simpopdf.com

مقایسه روش‌های طبقه‌بندی درختی جهت استخراج نقشه پوشش مرتعی با استفاده از داده‌های ماهواره‌ای (مطالعه موردی: حوزه آبخیز دویرج، استان ایلام)

مرزبان فرامرزی[الف]، حسن فتحی زاد[ب]، نسیبه پاکباز[ج]، بهزاد گل محمدی[د]

[الف] استادیار، گروه مرتع و آبخیزداری، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران (نگارنده مسئول)، پست الکترونیک: faramarzi.marzban@gmail.com

[ب] دانش آموخته‌ی کارشناسی ارشد بیابان‌زدایی، دانشگاه ایلام، ایلام، ایران

[ج] دانش آموخته‌ی کارشناسی ارشد زراعت، دانشگاه ایلام، ایلام، ایران

[د] دانش آموخته‌ی کارشناسی ارشد مرتعداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، ایران

چکیده. یکی از کاربردهای عمده داده‌های ماهواره‌ای طبقه‌بندی پوشش سطح زمین می‌باشد. طی سال‌های اخیر تعدادی الگوریتم طبقه‌بندی برای طبقه‌بندی داده‌های سنجش از دور ابداع شده‌اند. یکی از آن‌ها طبقه‌بندی‌های درختی می‌باشد. هدف اصلی این مطالعه مقایسه سه الگوریتم انشعاب روش طبقه‌بندی درختی (جینی، آنتروپی و نسبت بهره) برای طبقه‌بندی پوشش سطح زمین منطقه دویرج در استان ایلام می‌باشد. برای این، ابتدا تصحیحات هندسی و رادیومتری بر روی داده‌های ETM^+ سال ۲۰۰۷ صورت گرفت. سپس با بازدیدهای میدانی، طبقات مختلف کاربری اراضی تعریف و نمونه‌های آموزشی انتخاب گردید. نتایج حاصل از ارزیابی دقت تصاویر طبقه‌بندی شده نشان داد که روش طبقه‌بندی انشعاب جینی با ضریب کاپای ۸۹/۹۸ و دقت کل ۹۱/۱۷ ٪ داری بالاترین دقت بودند و بعد از آن طبقه‌بندی انشعاب نسبت بهره و همچنین آنتروپی با ضریب کاپای به ترتیب ۸۸/۴۵ و ۹۰/۶۵ و دقت کل ۸۶/۲۱ و ۸۶/۱۵٪ در مرتبه های بعدی قرار گرفتند.

کلمات کلیدی: طبقه‌بندی درختی، انشعاب جینی، انشعاب آنتروپی، انشعاب نسبت بهره، منطقه دویرج ایلام