Jurnal Teknologi

Full Paper

CLASSIFICATION METHODS FOR REMOTELY SENSED DATA: LAND USE AND LAND COVER CLASSIFICATION USING VARIOUS COMBINATIONS OF BANDS

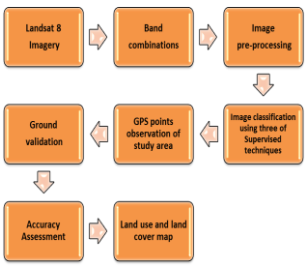
Nur Anis Mahmon[a*], Norsuzila Ya'acob[a], Azita Laily Yusof[a], Jasmee Jaafar[b]

[a]Wireless Communication Technology (WiCoT), Faculty of Electrical Engineering, Universiti Teknologi MARA, 404500 Shah Alam Selangor, Malaysia
[b]Faculty of Architecture Planning and Surveying, Universiti Teknologi MARA, 40500 Shah Alam, Selangor, Malaysia

Article history
Received
16 January 2015
Received in revised form
24 March 2015
Accepted
15 March 2015

*Corresponding author
nuranismahmon@yahoo.com

Graphical abstract



Abstract

Land use and land cover (LU/LC) classification of remotely sensed data is an important field of research by which it is commonly used in remote sensing applications. In this study, the different types of classification techniques were compared using different RGB band combinations for classifying several satellite images of some parts of Selangor, Malaysia. For this objective, the classification was made using Landsat 8 satellite images and the Erdas Imagine software as the image processing package. From the classification output, the accuracy assessment and kappa statistic were evaluated to get the most accurate classifier. Optimal performance was identified by validating the classification results with ground truth data. From the results of the classified images, the Maximum Likelihood technique (overall accuracy 82.5%) was the highest and most applicable for satellite image classifications as compared with Mahalanobis Distance and Minimum Distance. Whereas for land use and land cover mapping, the RGB 4, 3, 2 band combinations were found to be more reliable. An accurate classification can produce a correct LU/LC map that can be used for various purposes.

Keywords: Classification methods, land use and land cover, band combinations, accuracy

Abstrak

Pengkelasan 'Land Use' dan 'Land Cover' dengan menggunakan data satelit adalah salah satu kajian yang penting dan biasanya digunakan dalam aplikasi Remote Sensing. Berdasarkan kajian ini, pelbagai cara untuk mengkelaskan gambar akan dikaji dengan menggunakan combinasi RGB band yang berbeza. Kajian ini telah dijalankan di segelintir daerah di Selangor, Malaysia. Untuk mencapai objektif kajian. Data yang digunakan adalah gambar satelit Landsat 8 dan diproses dengan menggunakan perisian Erdas Imagine. Daripada keputusan pengkelasan, 'accuracy assessment' dan 'kappa statistic'akan dikira untuk mendapatkan pengkelasan yang paling tepat. Ketepatan pengkelasan akan dinilai dengan membezakan data dengan data di dataran. Berdasarkan keputusan yang diperolehi, 'Maximum Likelihood' menghasilkan ketepatan yang paling tinggi berbanding dengan 'Mahalanobis Distance' dan 'Minimum Distance'. Manakala untuk peta 'Land use' dan 'Land cover' combinasi band 4 3 2 adalah combinasi yang paling sesuai untuk menjalankan tujuan ini. Pengkelasan yang tepat akan menghasilkan peta LU/LC yang tepat dimana boleh digunakan untuk pelbagai tujuan.

Kata kunci: Cara pengkelasan, 'land use' dan 'land cover', combinasi band, ketepatan

1.0 INTRODUCTION

Remote sensing applications are surveying and recording activities of objects on Earth's surfaces. Satellite images have become the main source of remotely sensed data to obtain information for a variety of applications such as monitoring land use changes, forestry, agricultural, etc. [1]. In land use and land cover planning and management, remote sensing from satellite images is a very helpful tool for the identification and classification of land surface features [2]. Image classification is the most important part of remote sensing, whereby it is based on image analyses and pattern recognition [3]. In some cases, the classification itself can be the object of analysis. For example, land-use classification of remote sensing data can be used to produce maps for image analysis of the final product. Image classification has become an important tool for digital image classifications. The classifier loosely refers to computer programs that perform specific procedures for the classification of images. During this time, different image classification procedures have been used for different purposes by various researchers [4]. The objectives of this study are to test and evaluate the classification techniques to classify land use and land cover of several areas in Selangor, Malaysia using three types of classification techniques: Maximum Likelihood classifier, Mahalanobis Distance classifier and Minimum Distance classifier. Selection of the best band combinations in image classification is also an important part of the study because even if all bands are used, it still may not provide the best results in accuracy [5]. Therefore, this study also examines the optimum band selections for land use and land cover classification through comparison of the classification accuracies of various bands combinations.

1.1 Remotely Sensed Data

In this study, the Landsat 8 satellite sensor will be used to analyze land use and land cover classifications. The Landsat program has always been used to provide multi-spectral imagery of the Earth's land areas of moderate to high resolution (28.50 - 90.00 meter horizontal resolution) to support resource assessments, land-cover mappings as well as to track inter-annual changes in the environment. The Landsat sensor collects spectral information from the Earth's surface, besides having the ability to assess changes occurring on it.

1.2 Land Use and Land Cover

Land cover refers to the biophysical state of the Earth's surfaces and immediate subsurfaces including soil, topography, surface and groundwater and also human structures. Examples of land covers include forests, grasslands, croplands, wetlands and urban structures. Meanwhile, land use refers to the human use of land that involves management and modifications of natural environment for urbanization. Therefore, land use and land cover both defines natural and manmade coverings on the Earth's surface. The relationship between land use and land cover is not always direct and apparent. A single category of land cover may support many uses, while a single land use can involve the maintenance of several distinct land covers [6].

1.3 RGB Band Combinations

Individual bands can be composited from Red, Green and Blue (RGB) combinations in order to visualize the data in color. There are many different combinations that can be made as each has their own advantages and disadvantages [7]. Landsat 8 is composed of eleven different bands; each representing a different portion of the electromagnetic spectrum. For instance, band 1 represents coastal erosion, band 2 shows blue, band 3 shows green, band 4 represents red, band 5 shows Near Infrared (NIR), band 6 represents Short-wave infrared (SWIR 1), band 7 shows Short-wave infrared (SWIR2) [8], etc. Researchers [9] have stated that the combination of different bands in a satellite has issues in achieving good classification results. Therefore, to acquire the most accurate results for satellite image classification, the selection of suitable combinations of bands should be considered. An aspect for consideration in choosing the best combinations of bands is by selecting the bands according to the object on land and the type of classes which needs to be classified. This study uses some common Landsat RGB band combinations (color composites), which are: 4, 3, 2 RGB (natural color), 5, 4, 3 RGB (false color) and 7, 6, 4 RGB (infrared). These combinations were selected to evaluate their performance on which combinations can give the most accurate output for land use and land cover classification.

1.4 Image Classification

Image processing is the one technique that uses a computer to collect images for digital image manipulation. Image classification normally involves four steps:

Firstly is pre-processing of the image i.e. finding the band ratio, reduce haze, atmospheric corrections.

Secondly is the training sample, whereby it is the selection process of particular criteria or features that describes a pattern. Thirdly is the decision of selecting the most suitable technique for comparing targets with the image pattern. Lastly is assessing the accuracy of the image classification [10]. After the satellite image classification was done, land use and land cover mapping will be produced with an accurate classification.

2.0 EXPERIMENTAL

2.1 Study Area

The study area refers to the coverage area that is used in this study. Figure 1 shows the study area selected for this study; comprising of several districts in Selangor i.e. Klang, Petaling, Gombak and Hulu Langat. These areas were selected to cover the classification of land use and land cover. Remote sensing techniques will be used to classify the land use and land cover. A proper selection of the study area can assist in the selection of a correct satellite image.

Figure 1 Study area

2.2 Research Structure

Based on Figure 2, the methodology structure in this study is divided into three stages. The first stage is the satellite image processing, which includes the combinations of bands, image pre-processing process and image classification. The second stage is the data analysis, which analyzes the data with ground truth data to acquire the accuracy assessment for each classifier. This is followed by the third stage, which produces the land use and land cover map as the final output for image classification.

2.3 Combination of Bands and Image Pre-processing

The combination of three bands was determined using Erdas Imagine software. Findings from previous researchers have given a few combinations of the bands; for instance, a 3, 4, 7 band-combination is for detecting coastal or water boundaries. Bands 4, 3, 2 are used for vegetation and crop analysis, bands 4, 5, 3 for soil moisture and vegetation analysis, bands 3, 2, 1 for land cover and underwater features, bands 7, 4, 3 and bands 7, 4, 2 for change detection, soil type and vegetation stress [11, 12 and 13]. In addition, several published reports [14] also stated few band-combinations for Landsat 8. According to these, the band combinations 432, 543 and 764 have been selected for the classification of land use and land cover of the study area. Therefore, the best band combination is obtained by comparing the accuracy of the image classification.

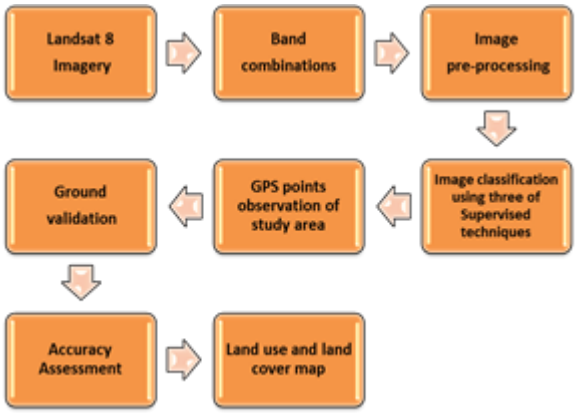


Figure 2 Research structure

Upon completing the combination of the three bands, the image proceeds with the image pre-processing using Erdas Imagine software. This stage is essential before the real processing and analysis of the image can be carried out. This stage involves image Geo-reference (geometric correction) including radiometric correction and image subset. Geo-reference image requires specific geodetic coordinate system (RSO coordinate system). Radiometric correction contributes to a clearer image without haze and lines. Pre-processing is often done to ensure the quality of the classification results.

2.4 Image Classifications

In this study, supervised classification classifiers are used to classify the images for land use and land cover classification of the study area. The Maximum Likelihood, Mahalanobis classifier and Minimum Distance classifiers are used and explained in this paper.

Supervised classification technique using maximum likelihood algorithm is the most commonly and widely used method for land cover classification [15]. This classifier is based on Bayesian probability theory. Maximum Likelihood Classification is a statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability. The Maximum Likelihood classifier is considered to give more accurate results than Minimum distance classification, however it is

much slower due to extra computations. It was found that the Maximum likelihood method gave the best results and both Minimum distance and Mahalanobis distance methods had overestimated agricultural land and suburban areas [16].

The Mahalanobis distance classifier is similar to the Minimum Distance, except that the covariance matrix is used in the equation. This algorithm assumes that the histograms of the bands have normal distributions [17]. Covariance are figured in so that clusters that are highly varied will lead to similarly varied classes. For example, when classifying urban areas, typically a class where pixels vary widely are usually not a highly varied class [18], hence it is slower to compute the minimum distance. Mahalanobis distance is parametric—meaning that it relies heavily on the normal distribution of data in each input band.

The Minimum distance classifier is based on training site data. This classifier characterizes each class by its mean position on each band. Minimum distance classifier is highly recommended in all image classification applications. The classification is performed by placing a pixel in the class of the nearest mean. The minimum distance algorithm is also more attractive since it is a faster technique than the maximum likelihood classification.

2.4 Accuracy Assessment

The results of the classification depend on the accuracy assessment and Kappa coefficient values. The percentage of accuracy of the classification result for all classifiers was calculated by analyzing it with the confusion matrix, which is also called the error matrix. Beside this, there are several indicators that have been used to show the classification results such as overall accuracy, producer accuracy, user accuracy and Kappa coefficient value [19].

$$k = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}.x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}.x_{+i}} \quad (1)$$

Where:

X_{ii} = Sum of diagonal input of error matrix
X_{i+} = Sum of row I of error matrix
X_{+I} = Sum of column I of error matrix
N = No. of elements in error

Based on [20], producer accuracy was calculated by dividing the number of correct objects of a specific class with the actual number of reference data objects for that class. Meanwhile, user accuracy was determined by dividing the number of correct objects of a specific class by the total number of objects assigned to that class. To perform producer accuracy, the proportion of labelled objects in the reference data was informed correctly. User accuracy, however, quantifies the proportion of objects assigned to a specific class that agree with the objects in the reference data. User accuracy indicates the probability that a specifically labelled object also belongs to that specific class in reality. It can also show the commission errors.

3.0 RESULTS AND DISCUSSION

3.1 Combinations of Band

The combinations of bands can give different significances in improving the land use and land cover classification. Three sets of band combinations were selected for supervised classification and evaluations on their performance. Figure 3, 4 and 5 represent the RGB band combinations; Figure 3 shows RGB band 4, 3, 2, Figure 4 RGB band 5, 4, 3 and Figure 5 RGB band 7, 6, 4. Band combination 432 is the near infrared channel (band 4), whereby this combination gives clearer land water boundaries and makes apparent different types of vegetation. This is a popular band combination for Landsat MSS data since it does not have a mid-infrared band. From Figure 4, the band combination and order makes vegetation look red instead of green whereas urban areas look blue in color. This combination is very popular with remote sensing specialists—mostly due to historical reasons [21]. Meanwhile, the 764 band combination shows the basic color of RGB in which vegetation is shown in green whereas water, blue.

Figure 3 RGB band combination 4, 3, 2

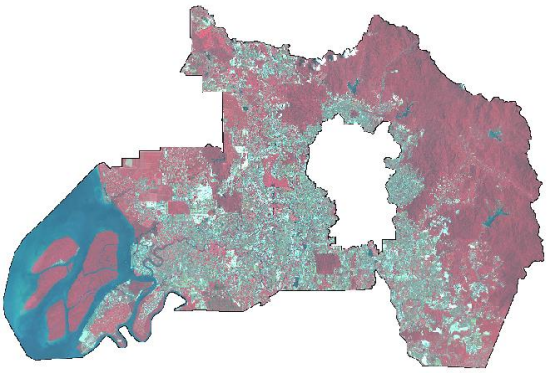


Figure 4 RGB band combination 5, 4, 3

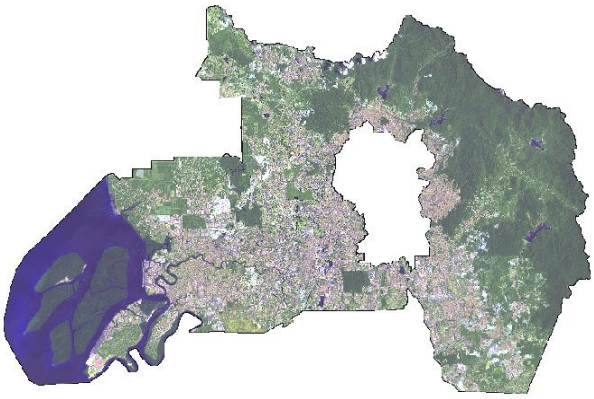


Figure 5 RGB band combination 7, 6, 4

3.2 Analyzing Different Classification Methods for Each Combination of Bands

In the classification stages, three methods of supervision were selected to classify the image. Two accuracies of classification output were tested in this study—namely, kappa coefficient and overall accuracy. The accuracy assessment was calculated by comparing the references data (validation point) with the output of the classification using Erdas Imagine software. The kappa coefficient of each classification is shown in Table 1. Band combination 4, 3, 2 shows the highest kappa statistic as compared with the other band combinations. From Figure 6, the Maximum Likelihood method gives the highest value of kappa statistic for all band combinations.

Table 1 Kappa coefficient for different types of classifiers

Classifier/ Band	4 3 2	5 4 3	7 6 4
Maximum Likelihood	0.8216	0.8010	0.7771
Mahalanobis Distance	0.5982	0.6964	0.6204
Minimum Distance	0.7893	0.7453	0.7384

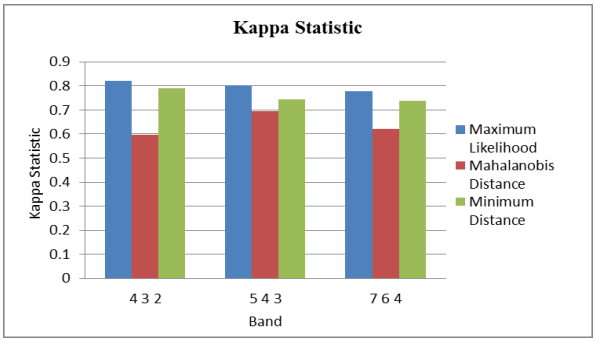


Figure 6 Comparison of Kappa Statistic for different types of classifiers for each band combination

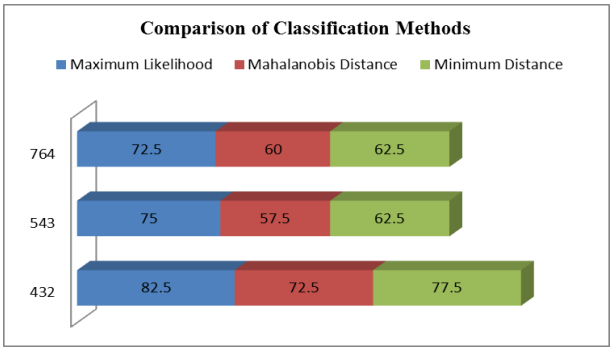


Figure 7 Comparison of Overall accuracy for different types of classifiers for all band combinations

Table 2 Classification efficiency of different methods for band 4 3 2

Techniques/ Class Name	Maximum Likelihood		Mahalanobis Distance		Minimum Distance	
	Producer Accuracy (%)	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*
Water bodies	80.00	100.00	33.33	100.00	80.00	100.00
Forests	50.00	66.67	50.00	100.00	57.14	80.00
Agriculture	25.00	100.00	60.22	45.50	75.00	75.00
Urban	87.50	100.00	87.50	100.00	92.86	72.22
Open land	100.00	77.27	100.00	61.54	50.00	100.00
Overall Classification Accuracy	82.50%		72.50%		77.50%	

Table 3 Classification efficiency of different methods for band 5 4 3

Techniques/ Class Name	Maximum Likelihood		Mahalanobis Distance		Minimum Distance	
	Producer Accuracy (%)	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*
Water bodies	66.67	100.00	16.67	100.00	80.00	80.00
Forests	50.00	100.00	25.00	100.00	50.00	37.50
Agriculture	40.00	50.00	40.00	40.00	40.00	25.00
Urban	95.00	76.00	90.00	62.07	69.23	75.00
Open land	50.00	100.00	20.00	100.00	50.00	100.00
Overall Classification Accuracy	75.00%		57.50%		62.50%	

Table 4 Classification efficiency of different methods for band 7 6 4

Techniques/ Class Name	Maximum Likelihood		Mahalanobis Distance		Minimum Distance	
	Producer Accuracy (%)	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*	*Producer Accuracy (%)*	*User Accuracy (%)*
Water bodies	66.67	100.00	57.14	100.00	57.14	100.00
Forests	25.00	100.00	14.29	100.00	40.00	33.33
Agriculture	40.00	66.67	90.00	55.67	50.00	25.00
Urban	100.00	66.67	94.44	56.67	86.67	76.47
Open land	40.00	100.00	50.00	100.00	28.57	100.00
Overall Classification Accuracy	72.50%		60.00%		62.50%	

In this section, the overall accuracy results of three methods for each band combination are shown in the Table 2, Table 3 and Table 4. The overall accuracy is expressed as the percentage of test pixels successfully assigned to the correct classes during classification. From the results, the Maximum Likelihood shows greater accuracy with overall accuracy of 82.50% for band combination 4, 3, 2, 75.00% for band combination 5, 4, 3 and then 72.50% for band combination 7, 6, 4. This method has the highest overall accuracy for all combinations followed by the Minimum Distance classifier. Concurrently, the Mahalanobis Distance classifier gives the lowest overall accuracy for all combinations of bands. The overall accuracy was obtained from different supervised classifiers. Therefore, according to Figure 7, the graphical results show the comparison of classification techniques in which the Maximum Likelihood techniques is always at the top of the graph for overall accuracies. In addition, no single method produced consistent accuracy for three sets of band combinations.

3.3 Analyzing Different Band Combinations

A wide set of band combinations were selected for supervised classification and performance evaluation. Different combinations produced different accuracy classifications. From the result, the Maximum Likelihood classifiers give the highest accuracy in comparison with the other techniques for all three combinations. Therefore, the Maximum Likelihood classifier results were selected to fulfil the objective, which is to identify the most suitable combination of bands for land use and land cover classification. Based on Figure 8, the accuracy obtained from 4, 3, 2 RGB band combination was the highest for all the classifications techniques. The mean value for the overall accuracy of this combination is 77.5%. In contrast, 5,4,3 and 7,6,4 RGB band combinations were the least accurate for classifying land used and land cover i.e. types water bodies, forest, agriculture, urban and open land in the study area.

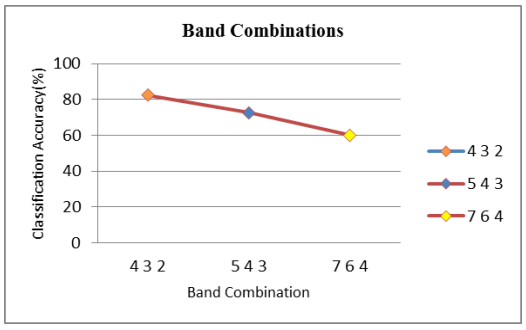


Figure 8 Comparison between three band combinations for overall classification of all the methods that were used in this study

3.4 Land Use and Land Cover Map

Land use and land cover maps play an important role in environment management. Five image classifications from the land use and land cover classifications can describe the environment of the study area. So, this project successfully achieved the objective for Land Use and Land Cover classification with the analysis of Landsat 8 satellite images. Figure 9 depicts the classified map as produced according to the best results of the image classification.

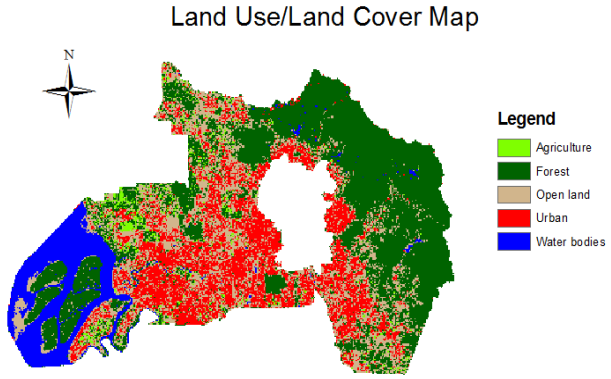


Figure 9 Land use and land cover map

4.0 CONCLUSION

The results presented in this study elaborated the efficiency of the three methods that have been used to classify the Land Use and Land Cover map of Selangor district. From these three classifiers, the Maximum Likelihood classification method produced the highest overall accuracy of 82.5%. As for the band combinations, the RGB 4, 3, 2 combination is the most suitable band combination for land use and land cover classification purposes. Moreover, five features from the image have been successfully classified, which includes Forests, Agriculture, Water bodies, Urban and Open Land. The high resolution images gave more detailed information of the classified map. In addition, the differences that have been analyzed in this study can act as guidelines for the community to choose the best technique with the most accurate combinations of bands in classifying land use and land cover. The classified images can also be used for planning and development purposes of Natural Resource Management in the future.

Acknowledgement

The authors would like to thank Agency Remote Sensing Malaysia (ARSM) for providing the satellite imagery. The author also wishes to thank UiTM Library Tun Abdul Razak Library for supplying topographic map of the study area.

References

[1] S. E. Franklin. 2001.*Remote Sensing for Sustainable Forest Management*. New York .Lewis Publishers,

[2] A. De Sherbinin. 2002. Land Use and Land Cover Change, 4 A Clesin Thematic Guides, Palisades, YN. Centre for International Earth Science Information Network of Columbia University. 67.

[3] Dean, A. M. and Smith, G. M. 2003. An Evaluation of Per–Parcel Land Cover Mapping Using Maximum Likelihood Class Probabilities. *International Journal of Remote Sensing*. 24(14): 2905-2920.

[4] Bruzzone, L., and S. B. Serpico. 2000. A Technique for Feature Selection in Multiclass Problems. *International Journal of Remote Sensing*. 21: 549-563.

[5] M. M. Elaalem *et al*. 2013. Performance of Supervised Classification for Mapping Land Cover and Land Use in Jeffara Plain of Libya. *International Conference on Food and Agricultural Sciences*. V55: 7.

[6] M. Seetha, I. V. Muralikrishna. Artificial Neural Network and Other Methods of Image Classification. *Journal of theoretical and Applied Information Technology*.

[7] Geospatial Innovation Facilities in Center for Biodiversity and Conservation.

[8] Kevin Butler. 2013. Band Combinations for Landsat 8. ArcGIS blog Esri.

[9] Jamshid Tamouk and etc. 2012. Satellite Image Classification Methods and Landsat 5TM bands. Department of Computer Engineering. EMU University, Famagusta, North Cyprus.

[10] Congcong Li *et al*. 2014. Comparison of Classification Algorithm and Training Sample Sizes in Urban Land Classification With Landsat Thematic Mapper Imagery. *ISSN Remote Sensing*. 968-983.

[11] Current Science. 2008. Discovery of Heavy Mineral-rich Sand Dunes Along the Orissa–Bengal Coast of India UsingRemote Sensing Techniques. *Current Science*. 94(8): 25.

[12] Davis, A and Allen, J. Landsat 7 at NASA Goddard SpaceFlight Center, 2012.

[13] M. Z and Abdullah, K. 2002. Evaluation, Conventional Digital Camera Scenes for Thematic Information Extraction. School of Physics Universiti Sains Malaysia, Penang, Malaysia.

[14] Kevin Butler. 2013. Band Combinations for Landsat 8. ArcGIS blog Esri.

[15] Jia, X. and Richards, J. A. 2006. *Remote Sensing Digital Image Analysis, An Introduction*. Taylor and Industrial Arts.

[16] Al-Ahmadi, F. S. and Hames, A. S. 2009. Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas. Kingdom of Saudi Arabia JKAU. *Earth Sci*. 20(1): 167-191.

[17] Perumal, K. and Bhaskaran, R. 2010. Supervised Classification Performance of Multispectral Images. *Journal Of Computing*. 2(2): arXiv:1002.4046v1.

[18] Swain, P. H. and Davis, S. M. 1978. *Remote Sensing: The Quantitative Approach*. US, McGraw-Hill, Inc,

[19] Sultan AlSultan. 2012. Innovative Satellite Image Map of R. Alkhabrta Area, Saudi Arabia Using High Resolution Image, School of Physics.

[20] Tso, B. and P. M. Mather. 2001. *Classification Methods for Remotely Sensed Data*. 1st Ed. Taylor and Francis, London.

[21] Ned Horning. 2004. Selecting the Appropriate Band Combination for an RGB Image Using Landsat Imagery. Center for Biodiversity and Conservation.