Comparative Assessment of the Unsupervised Land Use Classification by Using Proprietary GIS and Open Source Software

Muhammad Azizol Ismail*, Ahmad Nazri Muhamad Ludin and Nafisa Hosni
Department of Urban and Regional Planning, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

*Corresponding Author: azizolismail@gmail.com

Abstract. Mapping and investigating land use land cover (LULC) changes over a particular region is crucial for resource management, sustainability development, and holistic planning. An increasing rate of urban growth and urban sprawl could induce changes in land use as well as land transformation. However, accurate and up-to-date information about LULC is required for providing better understanding and assessing the environmental consequences of such changes. In this study, the 2017 image from the Sentinel-2A Satellite was utilized to demonstrate the land cover classification analysis in Iskandar Malaysia. Usually, land use classification analysis is conducted through proprietary GIS software. However, this decade shows the advancement in software development, thus the emerging of free/open source software in the geospatial world. Hence, to execute land cover analysis using the Unsupervised Classification technique, the proprietary GIS software (ArcGIS) and free/open source software (QGIS) were deployed. Then, the examination of accuracy assessment was carried out for the selected software. The sum of 250 random points was established for the assessment purpose. The results showed the overall accuracy for ArcGIS and QGIS were 82.80% and 80.40% respectively. The kappa coefficient for ArcGIS was 0.7395, while kappa coefficient for QGIS was 0.7094. Besides that, ArcGIS demonstrated better producer’s accuracy in the forest and agriculture land covers classification. Meanwhile, QGIS exhibited better producer’s accuracy in the built-up and water features classification. To summarize, ArcGIS and QGIS software are reliable to be used in the land cover classification.

1. Introduction
Urbanization is a common process that occurs around the globe as city is dynamic. The phenomena could induce significant changes on landscape pattern of a particular region. There is an increasing demand for innovative methods in implementing sustainable planning to examine current and past environmental condition [1]. In a given area, if the number of populations is increasing, the demands for the built-up area will also increase to occupy people needs. Unfortunately, other land covers (e.g. agriculture, forest, bare land) will decrease due to the changes of function to built-up. Accordingly, failure to keep monitoring the changes will lead to adverse impacts not only towards environment, but also its inhabitants. Therefore, establishing recent and accurate information about land use land cover (LULC) will give a better understanding and assessing the environmental changes efficiently [2]. This may be accomplished by computer-aided or visual analysis. Information about the land cover or thematic maps could be obtained through remotely sensed data classification. Remote Sensing is useful at
capturing the changes, and to extract the information change from satellite it demands an effective as well as automated change detection techniques [3].

Monitoring precise land cover is vital for decision makers, planners, and resource managers in implementing policies that related to sustainable natural resource management. Any changes to the land use would influence land cover, and conversely. Thus, land classification and mapping are considered as an integral approach to have an insight of eco-system.

The integration of satellite imagery with high spatial resolution, improved image processing approach, as well as Geographical Information System (GIS) has enabled the consistency of LULC monitoring activity [4]. Many works related to LULC analysis have been conducted by using proprietary software such as ArcGIS, ERDAS Imagine, and ENVI. This decade shows the emergence of Free and Open Source Software for Geospatial (FOSS4G) to enhance and empowering geospatial analysis.

There are basic principles that FOSS4G relies upon. These principles can be listed as freedom i) to run a software for any purposes, ii) to access, study, and modify the source code of software iii) to redistribute software either have been modified or not [5]. However, there is a lack of scientific studies on the use of FOSS4G in spatial planning [6].

Hence, this paper aims to investigate the capability of GIS and Remote Sensing tools in LULC classification, taking a part of Iskandar Malaysia (western development gate) as the case study. Specifically, we evaluated the capability of selected FOSS4G and ArcGIS software in performing LULC analysis.

2. Materials and Method

Figure 1 shows the flowchart of research. It starts with downloading satellite image, satellite images band combination, image classification using the Unsupervised Classification technique, and accuracy assessment. It is necessary to systematically compare the possibilities of deploying FOSS4G (QGIS) as an alternative to the established proprietary software (ArcGIS) in the field of sustainable spatial planning. The comparisons are conducted to investigate the overall accuracy, user’s accuracy, producer’s accuracy, as well as kappa statistic generated from both ArcGIS and QGIS software.

![Flowchart of study](image-url)
Study Area and Data
The study area is situated within the Western Gate Development of Iskandar Malaysia. The areas included such as Tanjung Pelepas, Tanjung Piai, Pulau Kukup, and part of Iskandar Puteri. There are three Ramsar sites available here namely Tanjung Piai Ramsar, Sungai Pulai Ramsar, and Pulau Kukup Ramsar. As noticed, mangrove areas need to be conserved from any harmful development as they are treasured with flora and fauna. The Sentinel-2 satellite image (year 2017) data was accessed via the United State Geological Survey (USGS) website.

Unsupervised Classification
Land cover features were extracted using an Unsupervised Classification. [7] defined the Unsupervised Classification as “a method by which pixels are assigned to spectral (or information) classes without the user having any prior knowledge of the existence or names of those classes”. This type of classification is generated based on the computed algorithm embedded within the GIS software. The algorithm groups pixels into spectral classes. It is most often executed through clustering method. The Isodata Clustering Algorithm is a well-known algorithm deployed in implementing Unsupervised Classification.

In QGIS, Unsupervised Classification was conducted using the Semi-Automatic Classification Plugin. Meanwhile, in ArcGIS it was performed with the ArcGIS Spatial Analyst extension through the Image Classification toolbar (Iso Cluster Unsupervised Classification). Then, the false color band composite of 4-3-2 (Red-Green-Blue) was selected. The satellite image categorized to the four (4) main classes; forest, agricultural area, built-up area, and water. Table 1 shows the class description for this study.

| Classes                | Description                                                                 |
|------------------------|-----------------------------------------------------------------------------|
| Forest                 | All types of forest such as wetland forest, highland forest                  |
| Agricultural area      | Type of agriculture including oil palm plantation, rubber                   |
| Built-up area          | Consist of activities including residential, industrial, transportation, road, built-up land, commercial and services |
| Water body             | All types of water such as river, lake, reservoir                           |

Accuracy Assessment
Accuracy of Remote Sensing image classification such as land covers could be evaluated by using the confusion matrix approach. Overall accuracy was derived from the sum of correctly classified pixels divided by the total number of pixels. Besides that, the accuracy for the individual class was calculated based on the user’s accuracy and producer’s accuracy approach. The user’s accuracy, or reliability indicates that predicted correct values belong to a class for an individual category. User’s accuracy was measured by dividing correctly classified pixels with classified total pixels. Meanwhile, producer’s accuracy reveals the probability that a value in a given class was correctly classified. The producer’s accuracy could be obtained by dividing correctly classified pixels with reference total pixels. Another measurement used in this study is the Kappa Coefficient. Kappa Coefficient denotes the agreement between classification and truth values. Perfect agreement is achieved if the value of Kappa is 1. Meanwhile, there is no agreement if the value of Kappa is 0. The outputs are evaluated based on the below assessment:

i) Overall accuracy (%) = (Correctly classified pixels / Total number of pixels)
ii) User’s accuracy / commission error (%) = (Correctly classified pixels / Classified total pixels)
   • Commission error = 1 – user’s accuracy

iii) Producer’s accuracy / omission error (%) = (Correctly classified pixels / Reference total pixels)
   • Omission error = 1 – producer’s accuracy

iv) Kappa coefficient (K)

\[
\kappa = \frac{N \sum_{i=1}^{n} m_{i,i} - \sum_{i=1}^{n} (G_i C_i)}{N^2 - \sum_{i=1}^{n} (G_i C_i)}
\]

Where:
- \( i \) is the class number
- \( N \) is the total number of classified values compared to truth values
- \( m_{i,i} \) is the number of values belonging to the truth class \( i \) that have also been classified as class \( i \) (i.e. values found along the diagonal of the confusion matrix)
- \( C_i \) is the total number of predicted values belonging to class \( i \)
- \( G_i \) is the total number of truth values belonging to class \( i \)

3. Results and Discussion

Once the Unsupervised Classification was conducted in both ArcGIS and QGIS, then the accuracy assessment should be conducted to examine how well the performance of each software’s classification. In this study, accuracy assessment was defined as a comparison of classification produced by each software (ArcGIS and QGIS) with the map from the Sentinel-2 Satellite image. 250 random points sampling approach was created for the assessment purpose. Figure 2 and Figure 3 demonstrated the output of the Unsupervised Classification by using ArcGIS and QGIS software. The Kappa statistic could be measured as follows:

- if \( K \leq 0 \), no agreement;
- 0.01-0.20, none to slight agreement;
- 0.21-0.40, fair agreement;
- 0.41-0.60, moderate agreement;
- 0.61-0.80, good agreement; and
- 0.81-1.00 as very good agreement

![Figure 2. Unsupervised classification in ArcGIS.](image)
For ArcGIS, the accuracy assessment derived from executing the Unsupervised Classification approach indicates an overall accuracy was 82.80% with the Kappa Statistic of 73.95%. These results indicated a very good agreement between the Unsupervised Classification map generated by ArcGIS software with the reference data. A Kappa value of 73.95% can be interpreted as 73.95% better classification would be expected by random assignment of classes. Table 2 shows the confusion matrix of the Unsupervised Classification in ArcGIS.

On the other hand, the overall accuracy for QGIS was 80.40% with the Kappa statistic of 70.94%. Meaning that, only 29.06% of the pixels’ classification were erroneous which consider as small and accepted. Like ArcGIS, QGIS also demonstrates a good agreement between software generated by the Unsupervised Classification map with the reference data. Table 3 outlined the confusion matrix of the Unsupervised Classification in QGIS.

Both Kappa statistics in ArcGIS and QGIS were equivalent which in the range of a good agreement. Thus, the results generated by ArcGIS and QGIS could be considered as high reliability. Both software exhibit good result in classifying forest and agriculture classes. These classes were identified most easily because if the large difference in spectral signature compared to the other classes. ArcGIS has higher producer’s accuracy for agriculture (90.99%) and forest (75.68%) compared to QGIS with the value 81.98% for agriculture and 71.62% for forest. However, QGIS demonstrates higher producer’s accuracy for built-up and water classification with the value of 86.79% and 91.67% respectively compared to ArcGIS with the value of 79.23% for built-up and 66.67% for water. Table 4 summarizes the comparison of the overall accuracy and Kappa statistic of the Unsupervised Classification technique in ArcGIS and QGIS.
a) Accuracy assessment using ArcGIS:

Table 2. Confusion matrix of unsupervised classification in ArcGIS.

| Class      | Agriculture | Built-up | Forest | Water | Row sum | User’s accuracy (%) |
|------------|-------------|----------|--------|-------|---------|---------------------|
| Agriculture| 101         | 2        | 8      | 0     | 111     | 82.79               |
| Built-up   | 10          | 42       | 1      | 0     | 53      | 85.71               |
| Forest     | 10          | 42       | 56     | 4     | 74      | 83.58               |
| Water      | 1           | 1        | 2      | 8     | 12      | 66.67               |
| Column sum | 122         | 49       | 67     | 12    | 250     |                     |
| Producer’s | 90.99       | 79.23    | 75.68  | 66.67 |         |                     |

i) Overall accuracy
   = 82.80%

ii) Kappa Coefficient
    = 73.95%

b) Accuracy assessment using QGIS

Table 3. Confusion matrix of unsupervised classification in QGIS.

| Class      | Agriculture | Built-up | Forest | Water | Row sum | User’s accuracy (%) |
|------------|-------------|----------|--------|-------|---------|---------------------|
| Agriculture| 91          | 5        | 13     | 2     | 111     | 84.26               |
| Built-up   | 3           | 46       | 4      | 0     | 53      | 83.64               |
| Forest     | 14          | 4        | 53     | 3     | 74      | 74.65               |
| Water      | 0           | 0        | 1      | 11    | 12      | 68.75               |
| Column sum | 61          | 34       | 47     | 8     | 250     |                     |
| Producer’s | 81.98       | 86.79    | 71.62  | 91.67 |         |                     |

i) Overall accuracy
   = 80.40%

ii) Kappa coefficient
    = 70.94%

Table 4. Comparison of the overall accuracy and Kappa statistic of Unsupervised Classification in ArcGIS and QGIS.

| Unsupervised classification | Overall accuracy (%) | Kappa accuracy (%) | Kappa coefficient |
|-----------------------------|----------------------|-------------------|-------------------|
| ArcGIS                      | 82.80                | 73.95             | 0.7395            |
| QGIS                        | 80.40                | 70.94             | 0.7094            |
4. Conclusion

Information about land classification is needed by many stakeholders for the decision-making process. With the spatial data, it can provide significant input for spatial planning and environmental protection purpose. The emergence of FOSS4G could make the geospatial analysis cost-effective yet powerful and promising. QGIS is one of the leading FOSS4G software. This study investigates the land cover analysis conducted in the selected proprietary GIS software (ArcGIS) and FOSS4G (QGIS). The result of this study found that ArcGIS shows higher overall accuracy which is 82.80% compared to QGIS which is 80.40%. However, the deficit is small. However, QGIS demonstrates better producer’s accuracy in classifying the built-up and water feature class. ArcGIS indicates the kappa coefficient of 0.7395 and 0.7094 for QGIS. To conclude, ArcGIS and QGIS are reliable software to be used in land cover classification analysis.

Acknowledgments
This work is financed by Zamalah/Institutional Scholarship provided by Universiti Teknologi Malaysia and the Ministry of Higher Education of Malaysia.

References
[1] Cavur M, Kemec S, Nabdel L, and Duzgun H S 2015 An evaluation of land use land cover (LULC) classification for urban applications with Quickbird and WorldView2 data in 2015 Joint Urban Remote Sensing Event (JURSE) IEEE pp 1-4
[2] Giri C, Zhu Z, and Reed B 2005 A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets Remote Sensing of Environ. 94 123-132
[3] Roy D P, P E Lewis and C O 2002 Justice Burned area mapping using multi-temporal moderate spatial resolution data—A bi-directional reflectance model-based expectation approach Remote Sensing of Environ. 83 263-286
[4] Rawat J S and Kumar M 2015 Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India The Egyptian J. of Remote Sensing and Space Science 18 77-84
[5] Bouras C, Kokkinos V, and Tseliou G 2013 Methodology for Public Administrators for selecting between open source and proprietary software Telematics and Informatics 30 100-110
[6] Friedrich C 2014 Comparison of ArcGIS and QGIS for applications in sustainable spatial planning (Doctoral dissertation, uniwien)
[7] John AR 2012 Remote Sensing Digital Image Analysis: An Introduction 5th Edition (Australia: Springer Science & Business Media) p 94