GUIDELINES FOR AUTHORS
ASIAN CONFERENCE ON REMOTE SENSING ACRS 2021

Kohei Cho[1] and Kyaw Sann Oo[2]
[1]Tokai University, 2-28-4, Tomigaya, Shibuya-k, Tokyo 151-0063, Japan,
Email: kohei.cho@tokai-u.jp
[2]Myanmar Peace Center, No. 11, Shwe Li Street, Kamayut Township, Yangon 11041, Myanmar,
Email: ksoo@myanmarpeace.org

KEY WORDS: Land use/Landcover, supervised classification, Google Earth Engine

ABSTRACT: Land use/landcover (LULC) change provides us insights on the human-environment interactions over time. These insights would be useful to come up with land management strategies as well as for future planning. In this study, the LULC change of Universiti Putra Malaysia (UPM) from year 2010 to 2020 is studied from Landsat-5 and Landsat-8 images. By leveraging Google Earth Engine (GEE), all the tedious satellite data preprocessing can be skipped as it is a cloud-based platform equipped with a multi petabyte analysis ready data catalogue. The classes being studied are water, road, urban, vegetation as well as barren land. The built-in machine learning algorithm, for supervised classification known as classification and regression tree (CART) was used to create classification maps for the years 2010, 2015 and 2020. The ground truthing was done by referring to historical images from Google Earth. The classification maps were assessed using confusion matrix. It was found that there were spectral similarities between road and urban which resulted certain misclassification. Next, the areas represented by each class was computed and compared from the year 2010-2015 and 2015-2020 to qualitatively detect the change. It was found that areas of roads and urban increased while barren land, vegetation and water decreased. However, in all three years vegetation covered the largest areas. Throughout the span of 10 years, vegetation covered most of the areas in UPM. The output of this research will aid the UI Green Metrics as it provides an objective assessment of the green areas of the campus. It will also help the university management as a tool in their tasks of better managing university facilities.

1. INTRODUCTION

1.1 Introduction

Land use and landcover (LULC) of a particular area of land evolves through time due to human-environment interactions. It is crucial to understand this process and evaluate the best land management strategies as well as future planning. Through Google Earth Engine (GEE), different land use categories can be classified by using built in machine learning algorithms (MLA) with simplified geospatial workflow as well as rapid prototyping and visualization of results Google Earth Engine (GEE) is a cloud-based platform accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) for highly sophisticated big data processing for remote sensing and Geographical Information System (GIS). It has enabled scientists, researchers, and developers to identify changes, track trends, and quantify variations on the Earth's surface by combining a multi-petabyte collection of satellite images and geospatial information with planetary-scale analytic capabilities. Currently Universiti Putra Malaysia (UPM) has an informative map which provides information about the buses to get around facilities and public amenities. Datasets on land use and landcover change is significant to support decisions prior to facilitating campus sustainability and effective environmental management such as various eco-friendly and effective learning activities, research, co-curriculum, and quality management system. Moreover, the percentages of green space (classified as vegetation) as well as other parameters like water, urban as well as roads must be kept track as UPM is ranked first in Malaysia and top 28th in the world in 2021 for the UI-Green metric World University Ranking. The evaluation for UI-Green metric is done via questionnaires and picture evidence. However, additional supporting quantitative evidence are needed to validate the responses. This study focuses on the detection of land use/ land cover change of Universiti Putra Malaysia (UPM), Serdang campus by comparing classification maps from years 2010, 2015 and 2020 produced using the classification and regression tree (CART) classifier algorithm built in in GEE and quantifying the variations in terms of area per class found from the classification maps.

2. METHODOLOGY

2.1 Study Area and Data Collection

Universiti Putra Malaysia (UPM) is one of the top 5 research universities in Malaysia. This university has two different campuses and in this study the area of interest is the main campus is in Serdang, Malaysia. The University, which began as the School of Agriculture in 1931, now boasts outstanding modern facilities and a vibrant approach to teaching and research, as well as a rich history of excellent services and achievements. By considering the type of sensor data needed, spatial resolution and acquisition period, images from Landsat-5 and Landsat-8 were selected. For this study all the bands will be used except the thermal infrared bands which are band 10 and band 11 of Landsat 8 and band 6 of Landsat 5. All these data will be utilized from the Google Earth Engine repository of geospatial datasets which is inclusive of observations from satellite which are available publicly. Landsat data is available in Earth Engine in its raw form, as Surface Reflectance, TOA-corrected reflectance, and in various ready-to-use computed products such as NDVI and EVI vegetation indices. The ground truthing was done by referring to the aerial image of UPM as well historical images from Google Earth. Landsat images used in this study are simple cloud free composites which are computed using various processing methods in Earth Engine. This method eliminates the impact of cloud cover on the acquisitioned image. The range of data acquisition was set from 1st of January till 31st December for all the three years being studied. A Landsat TOA composite is produced from a collection of Landsat scenes for the years 2010, 2015 and 2020 using ee.Algorithms.Landsat.simpleComposite. The Landsat scenes are imported from the data catalogue by using ee.Image Collection.

2.2 Research Framework

Contrary to the conventional procedure of data acquisition, processing, and analysis to produce the expected outcome, a simplified generic framework consolidated in one system was developed and introduced to better understand land use/landcover changes. In this study, supervised classification is carried out with the aid of Earth Engine (EE) API method; ee.Classfier. The classifier package includes many traditional machine learning methods. According to (Wahap, 2020), the CART algorithm showed the best accuracy compared to Random Forest (RF) and Support Vector Machine (SVM). Therefore, for this study the CART algorithm is implemented to produce three different classification maps. Post-classification, quantitative analysis in terms of areas covered by each class was done to further understand the effect of human-environmental change that has taken place in the area of interest (AOI).

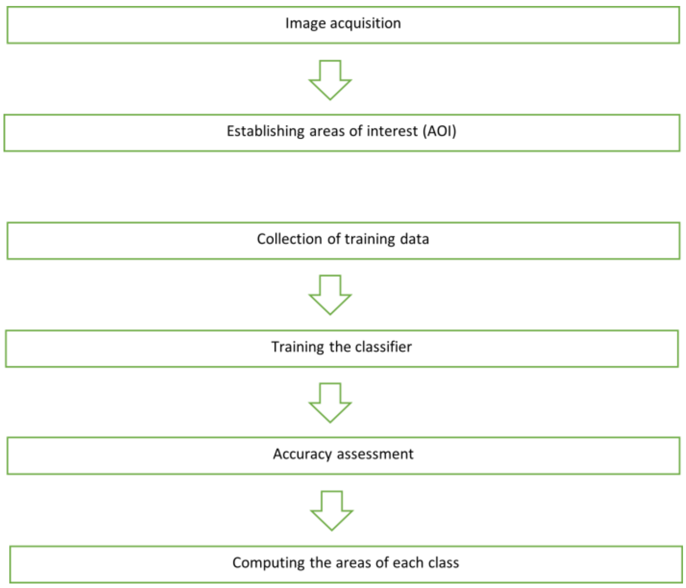

Diagram 1. The research framework

2.3 Training data collection and training the classifier

Prior to classification, training data are needed to be used as representative sample of reflectance spectra for each landcover class of interest. The machine learning model or classifier is highly dependent of these training points. Sufficient training data allows the classifier to accurately produce classification maps. A total of 1250 points were randomly marked with the aid of cloud free scene and the historical image from Google Earth that was produced for the year 2010, 2015 and 2020. The equal distribution of the training points will prevent overfitting of the model. This training points are then merged into a single Feature Collection. Next, by using ee.Image.sampleRegions, the pixels that are marked as Feature Collection will have one property per band of the input image, which simply means reflectance data for each point, from every band has been extracted. The collected samples were then trained using the CART algorithm to predict four simple classes which were defined earlier. The basis of the training are the spectral bands that have been extracted. Simply put, the algorithm then categorizes the spectral bands with similarities into one class. Using randomColumn() method. The training data was split into a 80:20 ratio, where at least 80% of the training points was used for training and more than 20% for validation.

2.3 Accuracy assessment

There are two different accuracy assessments here for the model which are overall training assessment as well as validation accuracy. The training accuracy illustrates the reliability of the model based on the examples that it was constructed. On the other hand, the validation accuracy dictates how well the model does when a different set of unknown examples are fed into it. When it comes to land use landcover maps, confusion matrix is an important element whereby the actual class is tabulated against the predicted class. Misclassified data can be identified through this method. In this study, the training and validation accuracy as well as the confusion matrix was used to assess the classification. From this confusion matrix, the user's accuracy (UA) and the producer's accuracy (PA) were also assessed. The user's accuracy is defined as the accuracy from the perspective of the map user/viewer. It basically shows how often the class on the map will be presented on the ground. On the other hand, the producer's accuracy is the perspective of the map maker on how often a particular class is shown correctly based on the ground feature. Furthermore, the F1 score or the harmonic mean of the recall and precision was also calculated to determine the test accuracy.

2.3 Area Computation

The areas of each class were found using ee.Image.pixelArea where the value of each pixel becomes the area in square meters. Then statistical pixel values of the image were reduced to the geometry which is the AOI in this case. This is done by using reduceRegion().The data was then represented in a pie chart. A vegetation map was also created to visualize the areas that has been covered by vegetation only. To compare the change that has occurred for each class, a class specific map was also produced.

3. RESULTS

Four land use/landcover classes are being observed in this study which is water, urban, road, vegetation, and barren land. The urban and road class were aimed to differentiate non-buildings and buildings. Vegetation class has the greatest number of inclusions. Land dedicated for agriculture as well as other forested land were classified together. Unused land is classified as barren land.

Table 1. Land use and landcover class and description

Class Name	Description
Water	lakes, pond, swimming pool,
Urban	faculties, residential colleges, administrative building, hospitals, labs, library, halls
Road	parking lots, roads
Vegetation	agricultural land, forested land
Barren Land	Unused land, land cleared for agriculture

3.1 Classification maps

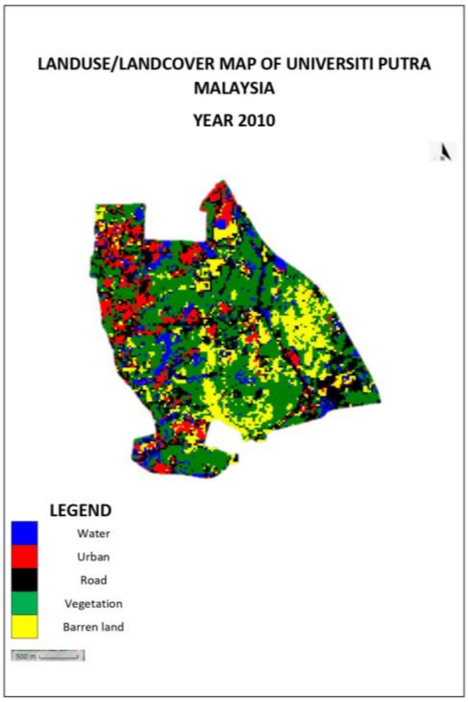

Figure 1. Classification map of UPM year 2010

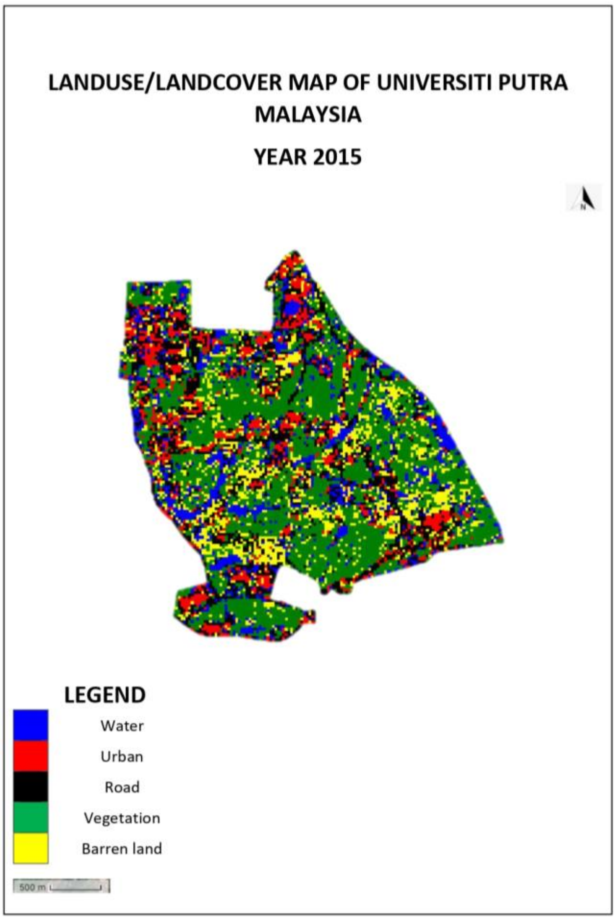

Figure 2. Classification map of UPM year 2015

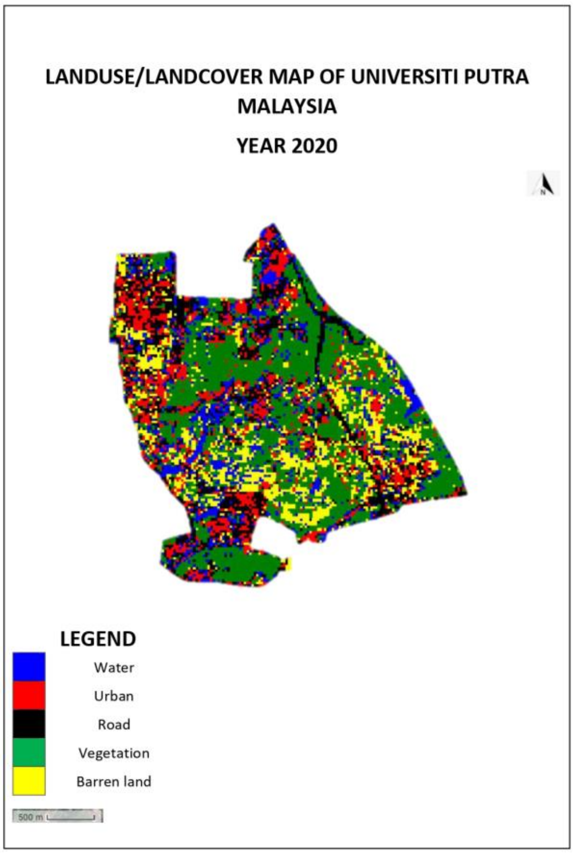

Figure 3. Classification map of UPM year 2020

3.2 Training and validation accuracy

The CART algorithm used showed a very good accuracy for the training data. For all 3 years, the training accuracy showed 99%. This is a good sign as it indicates that these is good model. The F1 score for the validation dataset shows an excellent score which translates to a good classification.

Class/Year	2010	2015	2020
Water	1	0.99	1
Urban	1	0.98	1
Road	1	0.99	1
Vegetation	1	1.00	1
Barren Land	1	1.00	1

3.3 Confusion matrix

To further asses the classification maps produced, confusion matrix or error matrix is used to visualize the performance of the algorithm. It consists of two dimensions which is inclusive of different combinations of predicted and actual values. The matrix can evaluate the actual target values with those that had been predicted by the model. The confusion matrix for the validation set is presented below.

To further asses the classification maps produced, confusion matrix or error matrix is used to visualize the performance of the algorithm. It consists of two dimensions which is inclusive of different combinations of predicted and actual values. The matrix can evaluate the actual target values with those that had been predicted by the model. The confusion matrix for the validation set is presented below.

Table 7. Confusion matrix for the year 2010, 2015 and 2020

2010						
Class Name	**Water**	**Urban**	**Road**	**Vegetation**	**Barren**	**Total**
Water	43	0	0	0	0	43
Urban	0	63	0	0	0	63
Road	0	0	52	0	0	52
Vegetation	0	0	0	46	0	46
Barren Land	0	0	0	0	49	49
Total	43	63	52	46	49	306
UA(%)	100.00	100.00	100.00	100.00	100.00	
PA(%)	100.00	100.00	100.00	100.00	100.00	

2015						
Class Name	**Water**	**Urban**	**Road**	**Vegetation**	**Barren**	**Total**
Water	52	0	0	0	0	52
Urban	1	41	0	0	0	42
Road	0	1	40	0	0	41
Vegetation	0	0	0	48	0	48
Barren Land	0	0	0	0	48	48
Total	53	42	40	48	48	306
UA(%)	98.11	97.62	100.00	100.00	100	
PA(%)	100.00	97.62	97.56	100.00	100	

2020						
Class Name	**Water**	**Urban**	**Road**	**Vegetation**	**Barren**	**Total**
Water	62	0	0	0	0	62
Urban	0	45	0	0	0	45
Road	0	0	44	0	0	44
Vegetation	0	0	0	50	0	50
Barren Land	0	0	0	0	48	48
Total	62	45	44	50	48	306
UA(%)	100.00	100.00	100.00	100.00	100.00	
PA(%)	100.00	100.00	100.00	100.00	100.00	

UA refers to user accuracy and PA refers to producer accuracy. Typically, these two accuracies are not the same. In UPM, there are many different materials that are used for the roof ranging from zinc, clay tile as well as asphalt compositions. Due to the presence of asphalt, there could been spectral similarities in between the two classes road and buildings. This was found by visually assessing the maps produced and ground truthing. Moreover, that the algorithm may have been confused itself with the spectral signatures of cars on the roads as the satellite images used are annual simple composites. There could have been cars on the specific sampled point which was denoted as road. This led to misclassifications in between buildings and road.

3.4 Area of Each Class

Statistical analysis in terms of area is done to observe the area covered by each class as well as to detect the change over the years in a statistical approach. The area represented by each class is presented in sq.km using a bar chart and the percentage is shown using a pie chart.

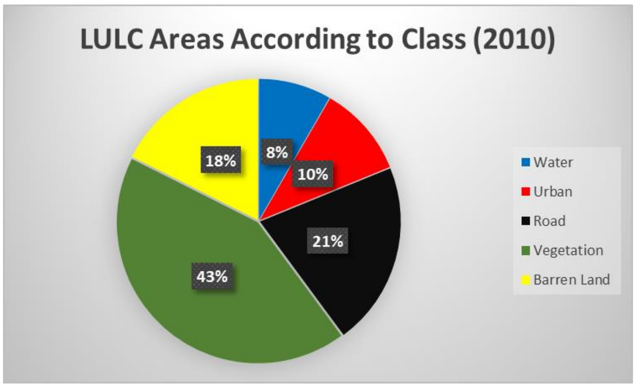

Figure 4. 2010 LULC distribution in percentage

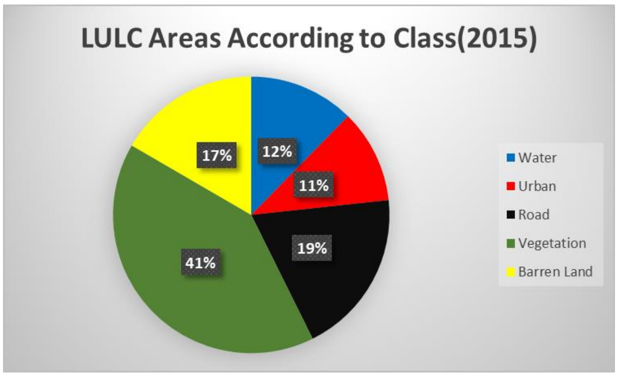

Figure 5. 2015 LULC distribution in percentage

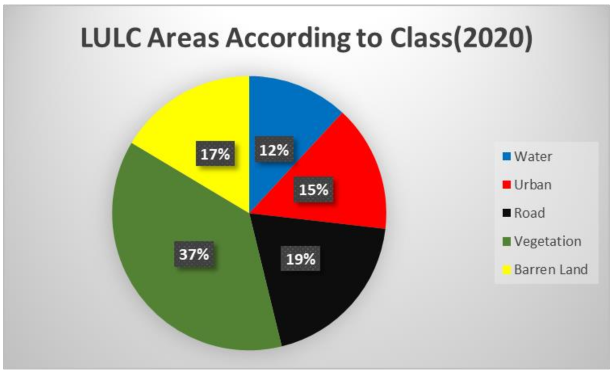

Figure 6. 2020 LULC distribution in percentage

From the data visualizations, it can be said that UPM is mostly covered by vegetation. Areas surrounded by barren land fluctuates as many agricultural activities take place in campus. Activities like land clearing takes place periodically. This also caused area of water to behave in the same way as barren land will be wet and cause puddles of water to be formed. Being an agriculture-based university, UPM facilitates eco-friendly learning activities to educate students on various agricultural processes such as land clearing, planting crops and more. As discussed earlier, the similarities in between the spectral signatures of roads and buildings causes the two land use classes to be best counted as one to avoid confusion. Over the years, UPM has undergone a steady development and the university management has made sure that the campus is always surrounded by lush greeneries ranging from landscapes, agricultural land as well as forests.

3.5 Change of landcover: water

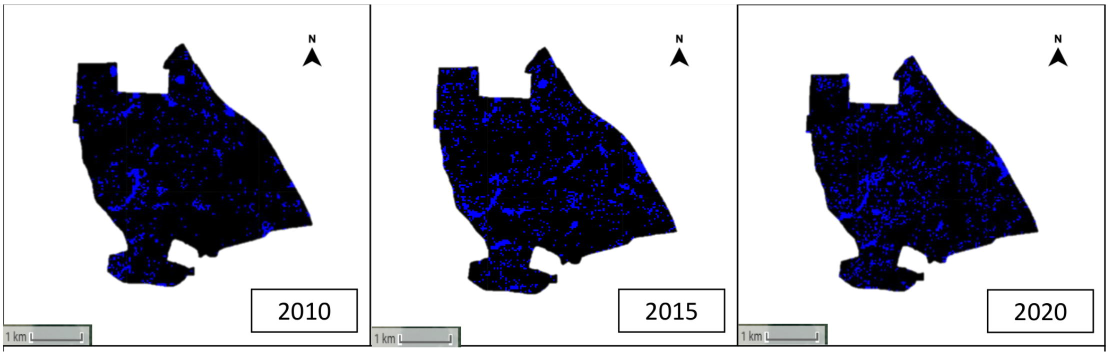

Figure 7. Change in landcover:water

From 2015 to 2020 there is reduction as a part of the ponds were covered with vegetation and the land cover changed. Figure below shows and evidence of a pond covered by grass thus reducing the depth of pond gradually and making it a grass covered land.

3.6 Change of landcover: vegetation

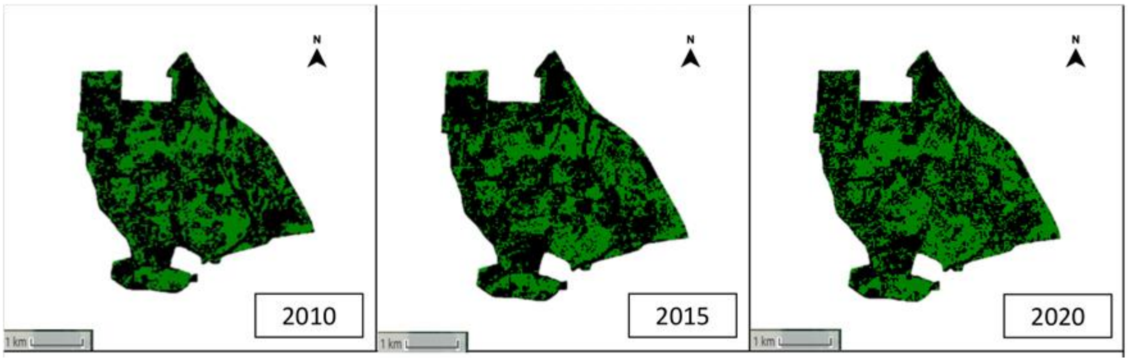

Figure 8. Change in landcover: vegetation

Vegetation is the most covered landcover for all three years. However, there is reduction of the areas covered over the years. The interesting part is over the time of 10 years trees matured and started producing lush green leaves which makes forests to appear denser. As discussed earlier, there are various agricultural activities that takes place in UPM. This is also a good effort to ensure the green areas are preserved and conserved. As for the year 2020, the Covid-19 pandemic has caused many activities to be halted. This has caused many existing agricultural activities to have reduced.

3.7 Change of landcover: barren land

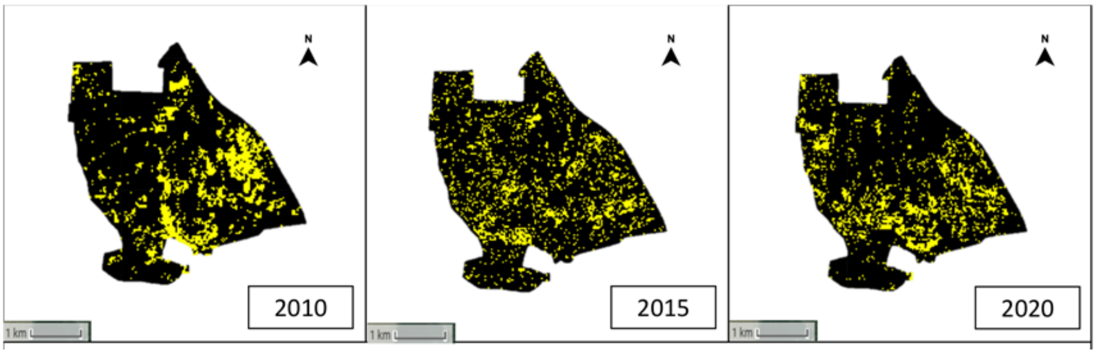

Figure 9. Change in landcover: barren land

Barren land reduced as much as 0.11sq.km from 2010 to 2015 and as much as 0.02sq.km from 2015 to 2020. The barren land reduced as many trees grew on the available unused land. The figure below illustrates a portion of land changed. Moreover, many cleared lands had buildings constructed.

3.7 Change of land use: urban

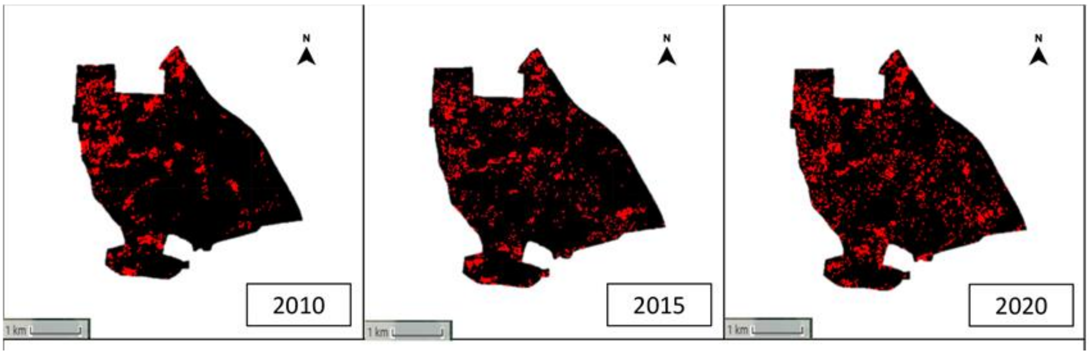

Figure 10. Change in land use: urban

Land use urban increased as much as 0.05sq.km from 2010 to 2015 and as much as 0.45sq.km from 2015 to 2020. This is due to emergence new buildings especially the new laboratory facilities in the faculty of engineering, Office of the Deputy Vice Chancellor (Research & Innovation) as well as University Putra Malaysia Teaching Hospital (HPUPM).

3.7 Change of land use: road

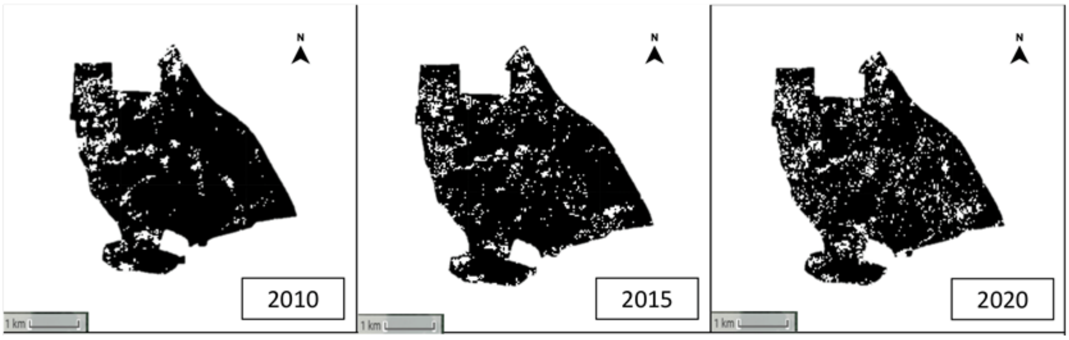

Figure 11. Change in land use: road

There are spectral similarities between road and building. As buildings increased the roads also increased, this is due to the parking lots as well as the roads surrounding the buildings The figure below shows a validation for claim above. University Putra Malaysia Teaching Hospital (HPUPM) has a large area of parking lots.

4. CONCLUSION AND RECOMMENDATION

4.1 Conclusion

The objectives of this study were achieved. The minor misclassifications were suspected due to Malaysia's location which results in high cloud cover in the acquired annual simple composite satellite images, rainfall as well as spectral similarities in between classes road and urban. From year 2010 to 2020, areas of roads and urban increased while barren land, vegetation and water decreased. However, in all three years vegetation covered the largest areas. UPM has to do more agricultural activities and even consider urban farming on the building roofs where applicable as well as indoor plants. GEE provided a hassle-free experience for a beginner to perform geospatial analysis as well to acquire remotely sensed images. High resolution images can be used so that it will have a finer resolution. Moreover, other algorithms can also be included to compare the results. Predictive LULC maps of the future can be created by using advanced unsupervised machine learning (ML) methods like deep learning. ML platforms like TensorFlow can be used for an increased performance of the algorithm by using deep learning algorithms like convoluted neural networks (CNN). However, it is a paid service if using GEE. Object based methods can be used to produce the training data as it is becoming a popular method to create classification maps. Besides that, auxiliary data like normalized difference vegetation index (NDVI) and Moderate Resolution Imaging Spectroradiometer (MODIS) can be used as validation data for improved classification. Since UPM has many types of crops, a crop specific classification would be useful as well.

5. REFERENCES

Wahap, N. A., & Shafri, H. Z. M. (2020). Utilization of Google Earth Engine (GEE) for land cover monitoring over Klang Valley, Malaysia. IOP Conference Series: Earth and Environmental Science, 540(1). https://doi.org/10.1088/1755-1315/540/1/012003