Utilization of Google Earth Engine (GEE) for land cover monitoring over Klang Valley, Malaysia

N.A Wahap¹,², Helmi Z.M. Shafri³

¹Malaysian Space Agency (MYSA), Space Technology Complex, 42700 Banting, Selangor, Malaysia
²Department of Civil Engineering, Faculty of Engineering, University of Putra Malaysia (UPM), 43400 UPM Serdang, Selangor.

helmi@upm.edu.my

Abstract. Geospatial Big Data is currently received overwhelming attention and are on highlight globally and Google Earth Engine (GEE) is currently the hot pot platform to cater big data processing for Remote Sensing and GIS. Currently few or no study regarding the usage of this platform to study land use/cover changes over years in Malaysia. The objective is to evaluate the feasibility of GEE as a free cloud-based platform by performing classification of Klang Valley area from Landsat composites of three different years (1988-2003-2018) using multiple Machine Learning Algorithms (MLA). The best classification results were then imported and further processed to quantify the changes over the years using commercial software. Although, the classification results are of high accuracy but CART shows the best accuracy with 94.71%, 97.72% and 96.57% in 1988, 2003 and 2018 in comparison with RF and SVM. Some misclassified pixels were encountered because the annual composited images were compiled without taken into considerations of crops phenological stages (paddy) which resulted to the misclassified agricultural land into urban and bare land. Hence, the selection and composition of data initially had to be structured and strategized prior to processing as they can affect the classification result and further analysis. Regardless, GEE has performed quite well and fast in term of time and processing complexity of multiple datasets with minimal human interaction and intervention. Generally, GEE has proven to be reliable in fulfilling the objectives of this study to evaluate the GEE feasibility by performing classification and quantifying the land use/cover of studied area and provide good base for further analysis using different platform.

1. Introduction

By the year 2030, it is expected that the urban population will surpass the rural and 53 percent forecasted was in Asia by 2050 [1] [2]. The population acceleration is generally due to the intensify population migration of rural and smaller town to the town generally in the developing countries for better quality of life and employment opportunities.

This phenomenon though seems normal but studies shown without proper planning and control, this population growth will bring various negative impact especially to the environment and natural surroundings which leads to declining and deprivation of infrastructures quality, agricultural and land use which resulted from micro-climatic changes and environmental pollution [3] [4]. The impact of
urbanization due to the industrialization has dynamic association between urban growth and environmental [5].

The current technology, allowing us to monitor and uses multiple resources and tools to evaluate and assess these impacts and scenario to our Earth and quality of life. With remote sensing of different platform and integration via satellite, plane, UAV, drone, and IoT (Internet of Things) technology, users will be provided with wider and extensive coverage and revisit capability to observe, combined with the higher resolution data and images which can offer users with an efficient way of monitoring land use/cover changes (LULC) over times. These information then can be processed using Geographical Information System (GIS) software to produce meaningful outputs and analysis required to serve decision support of multitude sector [6]. Monitoring land-use/cover changes (LUCC) over times, is one of the method to the imagery and visualize rapid urbanization while providing essential understanding of the land use/cover (LULC) dynamics over different spatial and temporal time scale [7][8][9].

Big Data is conventionally defined by volume, velocity and variety but that is not the case now as science and technology advance and develop in more rapid and drastically people need to be ready to face the consequences of increase in size, variety and update frequency. Hence, handling these data explosion requires high end computing techniques, modeling, architecture, methodologies, storage and solutions. Geospatial big data (GBD) refers to spatial data sets which surpasses the current capacity of computing systems and it is notable to mention that a big portion of big data in the real world is geospatial data, which kept growing rapidly by the least of 20% every year [10][11]. In today’s mapping systems, the challenge does not lie with the mapping technologies themselves, but rather, in the analytics platform that serves up the data.

Google Earth Engine is a free cloud-based platform for planetary-scale geospatial analysis which have supercomputing ability for complex calculation or massive processing problems [12]. GEE currently accessible to registered users through two web-based platforms: GEE Explorer and Code Editor. The GEE Explorer give access to user for a limited satellite imagery viewing, whereas with the GEE Code Editor, users were able perform analysis and customization by programming (JavaScript or Python) code. The Code Editor environment was built-in with mathematical and spatial operations in which users can use on the imagery separately or combined and tailored depending to the research goals.

A study in Thailand, used GEE to map the changes of a mangrove in three decades using NDVI and NDI derived from Landsat imageries for cloud masking and pixel-based classification of a Random Forest. This study however did the analysis in and out of GEE platform in order to prepare the top-of-atmosphere (TOA) data [13]. In Zambia, two approaches for land cover classification using the Landsat archive within GEE were used. A surface reflectance images were used and cloud, shadow and water pixel were removed using FMask algorithm. Random Forest classification was performed with (1) season-based composites, and (2) metric-based composites. These approaches were tested with Overall Accuracy exceeded over 85% in distinguishing cropland from non-cropland [14].

Goldblatt, Deininger & Hanson (2018), highlight the potential and accurate results of GEE to map the LULC in Ho Chi Minh City, Vietnam [15]. The study perform a supervised pixel-based Random Forest classification on optical and radar data, DMSP-OLS Nighttime Lights Time Series and Visible Infrared Imaging Radiometer Suite (VIIRS) to assess the relationship between the built-up extent and the population distribution. The fusion has successfully provide an accurate map of built-up and land use/cover.

A study in Turkey uses GEE to investigate the possibilities of identifying the changed areas within the selected areas. Data of Sentinel-1 and Multispectral Instrument (MSI) images of Sentinel-2 were used for change detection in the studied area. The 2015 and 2017 images (MSI and SAR) were fuse and the indices (Difference Built-up Index-NDBI, Bare Soil Index-BSI and Soil-adjusted Vegetation Index-SAVI) were utilized for extensive study on the changed areas. A binary supervised classification of Random Forest with BSI layer led higher accuracy of the classification. Strategically, LiDAR was recommended to be fused together to achieve higher accuracy [16].

Goldblatt, You, Hanson & Khandelwal, (2016), used imagery from Landsat 7 (annual TOA percentile composites) and Landsat 8 (raw images converted to TOA) as inputs by manually classified user identified
“built-up” or “not built-up” polygon using supervised image classification and detection of urban areas [17]. The Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) were introduce to improve the L7 classification. A pixel-based classification was apply to Classification and Regression Tree (CART), Support Vector Machines (SVM) and Random Forests (RF) gain high accuracy results with CART shows the largest improvement as the size of the training set increases.

The main objective of this paper is to evaluate Google Earth Engine (GEE) feasibility to study land use/cover changes (LUCC) over Klang Valley, Malaysia by determination of LUCC extent using multiple Machine Learning Algorithm (Classification and Regression Trees (CART), Random Forest (RF), and Support Vector Machine (SVM) in Google Earth Engine (GEE) platform. Subsequently, the best classification results were then use to quantify and map the temporal changes of the land use/cover over studied area from 1988 to 2018.

2. Study area
The area of interest in this thesis are consist of Selangor, Kuala Lumpur and Putrajaya to be named Klang Valley throughout this paper (Fig. 1). The Federal Territories of Kuala Lumpur and Putrajaya were enclaved inside Selangor, which were previously part of it. The state of Selangor has the largest economy and the most developed state in Malaysia in terms of gross domestic product (GDP) [18][19][20].

![Fig. 1 – study area location on map](image)

3. Data and methods
In order to perform a classification and change detection this paper will use two types of data and information (primary and secondary information). The primary information was gathered through the extracting and accessing the free and available Landsat series data from Google Earth Engine (GEE) platform.

The secondary information for this study was collected from several format. 2017 Selangor Landuse Map were obtained from Selangor Town and Country Planning (Selangor PLANMalaysia) and unstructured meet-up with key-individual of relevant field to study the relevant scopes of this study were use as references. Others references includes Google Earth images, various non-governmental organization, distributed and unpublished sources.
3.1. Primary information
The Landsat program is part of the USGS National Land Imaging (NLI) Program ensuring the continuity, reliable and comparable data since 1974. Landsat archives has started the open sourcing in 2008 and since then, the number of distribution has substantially increased from only 25,000 images in 2001 to 250,000 images per month by 2011 [21].

This paper uses annual composites Landsat Surface Reflectance images covering an area around 815784 hectares (~8,157 km²) of Selangor, Kuala Lumpur and Putrajaya for classification as based image for classification and further analysis. The reason is simply because Surface reflectance (SR) images provided in the GEE has already corrected the atmospheric effects (aerosol scattering and thin clouds) which improves detection and characterization of the surface changes. Three annual composites images from two Landsat Series were used in this paper (Table 1).

Table 1 - Landsat series Surface Reflectance data in respective years

| Data                              | Acquisition Year | Band      | Resolution (meter) |
|-----------------------------------|------------------|-----------|--------------------|
| Landsat 5 (MSS/TM)               | 1998             | Multispectral | 30                 |
| Landsat 5 (MSS/TM)               | 2003             | Multispectral | 30                 |
| Landsat 8 (OLI/TIRS)             | 2016             | Multispectral | 30                 |

The limitation of Surface Reflectance image in tropical and equatorial region would be the effectiveness of the correction where in this case reduced by the extensive cloud contamination. These can be seen in the image patches existed after the cloud removal procedure. In this case severely suffered by 2003 and 2018 composite image.

3.2. Secondary information
The 2017 Selangor Land Use Map were used and referred together in training and testing samples together with the primary data. The shapefiles of study area and other reference maps were obtained publicly from online platform and Selangor PLANMalaysia.

3.3. Methods
Various methods have been developed and introduced to measure land cover changes. In this paper, a consideration of generic setup was deployed to understands and evaluate the performance and feasibility of GEE a cloud-based platform to improve the conventional procedure of data acquisition and preparation to result handling and comprehensive analysis of remote sensing and GIS.

Land cover changes have been measured in many different ways whether through the land surface thermal changes, vegetation indices observation, post-classification and many more. However, these methods require extensive training and human assistance and intervention. A common framework for automated feature identification and to generalize objects for classification of different geographical are yet existed. This paper combines several generic methods to produce classifications maps in the GEE platform and afterwards to further analysed the results in the commercially available processing software.

The total analysis were projected through maps and graphs. Three temporal Landsat dataset of 1988, 2003 and 2018 were analysed individually. A simplification and generalization was injected in the methodology. The results of each classification of different years were analysed and change detection analysis was done by performing pre-and post-change detection techniques. The following chart represent the overall methodology performed in this study (Fig. 2).
3.4. **GEE image processing and manipulation**

Three customized GEE Code Editor scripts were used according to the individual year (1988, 2003, 2018) to avoid confusion and assist in processing time. The custom scripts developed in this thesis combine several components from official Google resources and other references. Each of the three scripts consists of five main components which includes:

a. Acquiring Landsat Surface Reflectance image collection
b. Study area and sampling selection
c. Supervised classification with MLA
d. Validation and assessment
e. Exporting results

3.5. **Acquiring Landsat Surface Reflectance image collection**
The first steps in accessing and acquiring a data started by a function to call and make a composite image by calling a stack or series of images from the Image Collections. Each of them has its own Image Collection and ID in GEE. In this case, the Landsat 5 and 8 SR collection were called and composited. An image collection from individual images or image merging from existing collections can also be derived in the GEE.

A filter function was scripted in to limit the acquired image to only in it chosen location and date of interest. As the images suffered from a severe cloud cover, a function to mask out the clouds using a pixel QA cloud band value provided within the SR products were utilized.

3.6. Study area and sampling selection
The image then was clipped to the user’s preference of designated shape (vector). There are four ways to obtain vector data in GEE whether uploading a shapefile directly to the personal Asset folder, or use an existing vector dataset in GEE, or draw manually or by importing an existing vector shape via Google Fusion Table. This study utilized importing the vector data from the Google Fusion Table into GEE platform.

3.7. Supervised classification with MLA
This study uses a pixel-based supervised classification with machine learning algorithm for land cover classification. An examination of the composite images was done to identify three sets of training and testing polygons (based on the images, Google Earth, and Selangor Landuse Map and Database) for six classes (Agricultural land, Forested land, Water bodies, Bare land, Urbanized land, Paddy field) for 1988, 2003, and 2018 as Feature Collection using the Geometry Tools and Import.

A presence of some sample selection biases existed as these samples selection is done manually based on the references available. The number of good samples was experimentally investigated by running the script repeatedly to gain an acceptable visual and statistical results.

The samples then were used to train the Classification and Regression Trees (CART), Random Forest (RF) and Support Vector Machine (SVM) classifier within Earth Engine platform (API Documentation). Subsequently, a map function was used to display the classification result. Extreme caution must be taken as the number of training and testing pixel and the complexity of the classifier will affect the processing time and results delivery.

3.8. Validation and assessment
In this paper, a ground truth data or the testing samples as explained in the previous subsection were collected by interpreting existing image, Google Earth of respective year and GIS database acquired from Selangor PLANMalaysia.

Confusion matrices was used to assess the accuracy of supervised classifiers and to get a true validation accuracy, a ‘testing’ data must be introduced and allocated to the classifier. In order to simplify and display how convenient it is to use GEE, the previous collected samples have been scripted to randomly segregate by 70:30 percent for training and testing. The script was written to holds out data for testing, then applies the classifier to the testing data and assesses the Confusion Matrix for this withheld validation data. The validation results then were printed in the Console on the right side of the Code Editor or can be exported into a table properties to the Google Drive.

3.9. Exporting results
Exporting results may serve many purpose as it provides a long term record of output gain from the analysis. The final step of GEE processing is to export relevant results from the analysis such as charts, images, maps, tables, etc.

3.10. Post-classification and analysis
This study focuses on the land cover change over a long period (1988–2018) of time on the study area. The overall scientific analysis determined the state, dynamics, and trend of changing landscape of Selangor and two enclaved Federal Territories were done outside GEE environment which explore the issues of compatibility and comparability between classification images and other software.

4. Results and analysis
A thorough analysis and discussion were made in a qualitative as well as quantitative manner. This chapter explored the feasibility and performance of GEE to enhance the execution of normal remote sensing and GIS processes.

4.1. Image quality and land cover classes
As described in previous chapter, the analysis made in this thesis were majorly based from Landsat-5 and Landsat-8 annual composite images and the quality of the Surface Reflectance composites images varies according to the intensities of the cloud covers over the chosen year. Taken into account that Malaysia is located in the equatorial region that suffers from cloud contamination throughout the year, it is expected that the quality of the derived images endures and inherit several flaws. This can be observed in all three images especially in 2003 and 2018 composite image where several image patches were found in the scene. Fig. 3 displayed and exported False Colour Composite image of respective years in GEE platform.

![Surface Reflectance image of Selangor for the year 1988, 2003 and 2018.](image)

By using the created training and testing polygons on the three images described in the previous chapter, the results of the classification were done and validated. Three classification maps for each year with three different machine learning algorithms were obtained. Based from the PLANMalaysia database and USGS Level 1 Classification Scheme, six (6) land cover classes were recognized as tabulated in Table 2 [22].
| Class Name      | Description                                                                 |
|-----------------|-----------------------------------------------------------------------------|
| Agricultural land | Cropland and pasture, orchards, groves, vineyards, nurseries, and ornamental horticultural areas, confined feeding operations, other agricultural land. |
| Forested land   | Deciduous forest land, evergreen forest land, mixed forest land.             |
| Water body      | Streams and canals, lakes, reservoirs, bays and estuaries.                  |
| Bare land       | Dry salt flats, beaches, sandy areas other than beaches, bare exposed rock, strip mines, quarries, and gravel pits, transitional areas, mixed barren land. |
| Urbanized land  | Residential, commercial and services, industrial, transportation, communications, and utilities, industrial and commercial complexes, mixed urban or built-up land, other urban or built-up land. |
| Paddy field     | Paddy field                                                                |
4.2. Classification results and output

Three (3) Machine Learning Algorithms consist of Classification and Regression Trees (CART), Random Forest (RF) and Support Vector Machine (SVM) were chosen for pixel-based supervised classification on 1988, 2003 and 2018 images (Fig. 4). The general results show, CART has consistently outperformed RF and SVM in classifying all the pixel into its classes in those three years (Table 3).
Fig. 4 - Classification of CART, RF and SVM by the year of (a) 1988; (b) 2003; and (c) 2018.

Table 3. CART land cover distribution over 1988, 2003, and 2018 in Hectares and Percentage

| Class Name | 1988          | 2003          | 2018          |
|------------|---------------|---------------|---------------|
|            | Area (Ha)     | %             | Area (Ha)     | %             | Area (Ha)     | %             |
| AG         | 399161.62     | 48.93         | 358819.63     | 43.99         | 327070.93     | 40.09         |
| FR         | 322775.97     | 39.57         | 286418.51     | 35.11         | 274807.78     | 33.69         |
| WT         | 4844.46       | 0.60          | 2284.02       | 0.28          | 3834.66       | 0.47          |
| BL         | 5902.35       | 0.72          | 16567.31      | 2.03          | 18635.99      | 2.28          |
| URB        | 42347.28      | 5.19          | 86185.50      | 10.57         | 124214.68     | 15.23         |
| PD         | 40752.99      | 4.99          | 65509.70      | 8.03          | 67220.63      | 8.24          |
| Total      | 815784.66     | 100           | 815784.66     | 100           | 815784.66     | 100           |

The first sets of analyses examine the performance of the three selected Machine Learning Algorithm visually and statistically in comparison with the composite images. Through these observations, CART has consistently displayed good accuracy in both visual and statistics results. Table 4 compares the overall accuracy and kappa coefficient of classified images of respective years and MLA.

Table 4 – overall accuracy and kappa coefficient of classifications and respective years

| Year | CART        | RF          | SVM         |
|------|-------------|-------------|-------------|
|      | OA (%)      | Kappa       | OA (%)      | Kappa       | OA (%)      | Kappa       |
| 1988 | 94.71       | 0.918       | 93.49       | 0.899       | 91.44       | 0.868       |
| 2003 | 94.72       | 0.92        | 94.26       | 0.913       | 89.03       | 0.830       |
| 2018 | 96.57       | 0.926       | 95.81       | 0.909       | 92.25       | 0.833       |

The classification results in this thesis were obtained at the level of pixel with spatial resolutions of 30m. For 1988 pixel-based supervised classification, CART display an undeniable performance with 94.71% overall accuracy and 0.918 for kappa coefficient. But this does not mean that RF and SVM does very badly. Both of the classifier produce generally good result visually and statically but CART has produce much higher accuracy as seen in Table 4, in the previous section.

The results exhibit that Agricultural land occupying the largest portion of the study area with 399,162 ha which is equal to ~49% of the study area (Figure 5(a)). This is mainly found in the central region and coastal area which is known of agricultural activities mostly in rubber and oil palm plantation. The next
land cover class with the highest area coverage is the Forested land with 322,776 ha about ~39% which is scattered more around the upper right portion of the study area. This area shared Titiwangsa Mountains which house the national reserve forest with its neighbouring states.

The Urbanized land consist of several developed areas and scattered all over Selangor and Kuala Lumpur. Around 42,347 ha or ~5% of the land were urbanized and developed in 1988 and mostly in Kuala Lumpur and Klang valley area. Selangor also play an important role in paddy planting and one of Malaysia major supplier of rice. In 1988, it is recorded around 40,753 ha (~5%) of paddy field were recognized in from the image mostly located in Sekinchan, Selangor. Both Bare land and Water body cover about 1% of the studied area which amounted to 5,902 and 4,844 hectares. Figure 5(b) and 5(c) also display the 1988 land cover distribution in area and percentage.

Fig. 5 – 1988 land use/cover distribution in (a) map; (b) hectares; (c) percentage
CART also exhibit the highest accuracy and statistical results for 2003 classification with cautious and extra attention has been put into the training and testing sample selection because the image quality of 94.72% overall accuracy and 0.920 for kappa coefficient.

2003 CART displayed Agricultural land was still inhabiting the largest portion of the study area but with lower value with 358,819 ha which is equal to ~44% of the study area (Figure 6(a)). The second highest area coverage is the Forested land with 286,419 ha covering about ~35% of the study area. The Urbanized land consist of several developed areas and scattered all over but with higher concentration in the west coast of Selangor and Kuala Lumpur city centre.

Around 86,186 ha or ~11% of the land were classified as urbanized and developed land. The classified paddy planting area was also manifest around 65,510 ha (~8%) in the classified map. Both Bare land and Water body cover remain around about 1% of the area with 7,452 and 6,609 hectares. Figure 6(b) and 6(c) shows the 1988 land cover distribution in area and percentage.
2018 classification also shown a similar trend in CART accuracy and kappa (κ) value with 96.57 overall accuracy and 0.926 in kappa the highest of the three years. Classification results recorded for RF and SVM also turn out to be the best among the three years. In the result (Figure 7(a)), a randomized pixel was classified as paddy field when in reality the pixel classified was not at all a paddy planting area. This might outline the disadvantage of having an annual composited images which disregard paddy planting and harvesting season and the training samples does not provide enough information about the sampled pixels. Nevertheless, the value is still somewhat comparable to the original data and references data. In 2018 (Figure 7(b) and 7(c)), agricultural land although still covers the most part of area (~40%) but the covered area has decreased to 327,071 ha. This declining features also happened to forested land which decreased to 274,808 ha (~34%). Whereas, all other land use/cover has increased in volume starting with water bodies to 3,834 ha (~1%), bare land increase to 18,635 ha (~2%), urbanized land inhabits over 124,215 ha (~15%), and paddy field to 67,221 ha (~8%).

![2018 CART Classification](image)

4.3. Validation and accuracy assessment

Fig. 7 – 2018 land use/cover distribution in (a) map; (b) hectares; (c) percentage
The next step in the methodology was accuracy assessment analysis of these data. The traditional analysis of accuracy assessment data begins with an error matrix, sometimes also called an agreement or confusion matrix [23]. An error matrix summarizes the correct classifications and misclassifications of CART classification of respective years in a contingency table format as highlighted in the Table 5.

From the observation, based on the recommended methods of random training and testing samples splitting methods, GEE will produce different sets of confusion matrix and accuracy which were dependent on the number of pixels includes and different time of day. Hence the validation and classification accuracy for random splitting were also dependent on the training and testing selection.

Table 5 – Confusion matrix of CART classification for 1988, 2003 and 2018

| Class Name | AG   | FR   | WT   | BL   | URB  | PD   | Total |
|------------|------|------|------|------|------|------|-------|
| 1988       |      |      |      |      |      |      |       |
| AG         | 5439 | 305  | 0    | 1    | 4    | 39   | 5788  |
| FR         | 391  | 7425 | 0    | 0    | 1    | 41   | 7858  |
| WT         | 0    | 0    | 115  | 0    | 0    | 2    | 117   |
| BL         | 0    | 0    | 64   | 0    | 0    | 64   |       |
| URB        | 32   | 0    | 0    | 3    | 298  | 3    | 336   |
| PD         | 59   | 19   | 1    | 0    | 1    | 2812 | 2892  |
| Total      | 5921 | 7749 | 116  | 68   | 304  | 17055|       |
| UA (%)     | 91.86| 95.82| 99.14| 94.12| 98.03| 97.07|       |
| PA (%)     | 93.97| 94.49| 98.29| 100  | 88.69| 97.23|       |

| Class Name | AG   | FR   | WT   | BL   | URB  | PD   | Total |
|------------|------|------|------|------|------|------|-------|
| 2003       |      |      |      |      |      |      |       |
| AG         | 10083| 280  | 0    | 2    | 22   | 278  | 10663 |
| FR         | 306  | 9037 | 0    | 0    | 50   | 9393 |       |
| WT         | 0    | 2    | 95   | 0    | 1    | 98   |       |
| BL         | 1    | 0    | 0    | 176  | 0    | 177  |       |
| URB        | 26   | 0    | 0    | 902  | 32   | 960  |       |
| PD         | 307  | 35   | 0    | 2    | 3766 | 4110 |       |
| Total      | 10723| 9354 | 95   | 176  | 926  | 4127 | 25401 |
| UA (%)     | 92.2 | 97.26| 100  | 98.85| 97.04| 93.48|       |
| PA (%)     | 96.12| 95.16| 94.44| 100  | 94.16| 88.66|       |

| Class Name | AG   | FR   | WT   | BL   | URB  | PD   | Total |
|------------|------|------|------|------|------|------|-------|
| 2018       |      |      |      |      |      |      |       |
| AG         | 1606 | 94   | 0    | 0    | 2    | 23   | 1725  |
| FR         | 144  | 8964 | 0    | 0    | 2    | 9110 |       |
| WT         | 0    | 0    | 41   | 0    | 1    | 42   |       |
| BL         | 0    | 0    | 0    | 131  | 5    | 0    | 136   |
| URB        | 5    | 1    | 0    | 5    | 354  | 86   | 451   |
| PD         | 23   | 5    | 0    | 43   | 1272 | 1343 |       |
| Total      | 1778 | 9064 | 41   | 136  | 404  | 1384 | 12807 |
| UA (%)     | 90.33| 98.9 | 100  | 96.32| 87.62| 91.91|       |
| PA (%)     | 93.1 | 98.4 | 97.62| 96.32| 78.49| 94.71|       |
4.4. Change detection and analysis

In reference with the result (Figure 8) gain from the classification of 15 years’ interval, we can see that green cover types showing a decreasing pattern and developing area display an increment in 1988 to 2003. This trend also can be seen in 2003 to 2018 where all the green cover types exhibit a decreased in value except for paddy field area.
Fig. 8 – land cover changes (a) 1988-2003; (b) 2003-2018; and (c) 1988-2018.
Figure 9, exhibit the overall changes (1988 – 2018) between land use/cover types. Table 6 tabulate the overall and 15 years’ interval land use/cover changes between classes. The results indicate an impressive changes happened throughout the thirty years of observation. In line with the hypothesis, the Urbanized and bare land has displayed rapid increased in area expansion in thirty years which doubled from 42,347 to 124,214 ha with total of 81,867 area expanded (urbanized land) whereas, bare land has gained over 12,734 ha which adds up to a total of 18,636 ha. Which proven a strong correlation between the urban and bare land expansion with the decremented agricultural land and forested land indicating the land use/cover transition for urbanization processes in the literature review.

Though in this study, the misclassified pixels (especially in 2003 and 2018 images) is suspected was inherited from the low quality of based images due to heavy cloud contamination which resulted to image patches. The misclassified pixels were also suspected comes from the phenological cycle or phase.
4.5. Misclassified pixels and crops phenological cycle

Despite a dominant inhibition of the paddy plantation area in upper right corner of the images is challenging to map using the annual composited images from Landsat data given its spectral similarity to various land use/cover of surrounding the area. This section addresses this challenges in three composited images of 2003, and 2018 acquired from the Landsat.

Figure 10, exhibit the 2003 and 2018 images up to 1:150,000 scale. The images display the intensity of patches (in red circle) of different hue and pattern compared to the same land use/cover type (forested land and paddy), in order to fill in the gap of clouds from a scene which resulted from a few of composited image of different month in a year.

![Fig. 10 – image patches intensity of composited images of 2003 (left) and 2018 (right)](image)

As shown in Figure 11, the spectral profile of respective land cover types indicating a similarity between paddy and agricultural land and bare land. This similarity was expected due to the fact that the image composition is of various phenological season of seeding, growing, harvesting and post-harvesting. This were proven by previous studies of mapping a seasonal cropland especially paddy and rice plantation which utilizing the phenological stages of the crops in data collection and analyses [24] [25].
Additionally, since the image were composited from multiple months in a year, some of the paddy plots were of a fallow land resulted to drought conditions hence a similarity in spectral value of the bare land. Whereas, some of the image patches were of growing stages and has a closer spectral proximity to the agricultural land.

![Spectral profile of paddy, urbanized land, agricultural land and bare land](image)

Fig. 11 – Spectral profile of paddy, urbanized land, agricultural land and bare land

Although reference data for 2018 were obtained from the authority were used in this study but it still does not assist the automated classification in GEE. This however can be improved with more intervention from the user but limiting the automation part or adding up to the complexity of processing in the GEE.

Ergo, classification of the images by using pixel-based method indicate a randomized paddy classification in the bare land, urbanized land, and agricultural land. This highlighted another limitation and constraint of using the GEE Code Editor using JavaScript was users has no knowledge and back-end information of activity behind the results which includes multiple complex processing.

5. Discussions and recommendations

As previously discussed, this paper is focus to only cover the temporal changes of Klang Valley (Selangor and Federal Territories of Kuala Lumpur and Putrajaya) with three sampled annual cloud-free images of 1988, 2003, and 2018. All data/images were acquired, processed, and analyzed in GEE platform is of Landsat imageries with 30m of spatial resolution which as we can see has affected various aspects in the processing methods.

The limitation in this aspect was, Malaysia is located in the equatorial region which suffers from heavy clouds and atmospheric constituent. Furthermore, an annual cloud-free masking has degraded the image composition results which later affected the pixel-based supervised classification. Hence, issue of random misclassified pixels of paddy field area.

The classification of images was done in multiple machine learning algorithm (MLA) and change detection was modelled based from the best classification results on different platform as at the time being, GEE does not provide post-classification comparison algorithm.
This paper also ventures in the study of large scale LUCC analysis on cloud-based platform to measures and quantify the changes and make a necessary judgment on its performance and how to manipulate the results into something meaningful. Overall, GEE is proven to be a great tool and alternatives to users without disregarding the commercially established remote sensing and GIS processing software and platform.

The classification results of the three machine learning algorithms are also quite satisfying and produced good accuracy with CART performed the best. Although, there was an issue with misclassified pixels due to the inconsideration of phenological stages of the crops but the overall capability of GEE to perform supervised classification is proven in term of other technical and processing aspects. It is also recommended for future works to ingest the object-based (OBIA) classification and spectral indices fusion into different phenological seasons for better accuracy and not just depending on the statistical validation.

Regardless, GEE has effortlessly performed multiple function in preparing users with datasets. Within seconds, users can easily acquire a complete clipped, mosaicked, composite image of chosen spatial location. It was hassle-free, no messy codes or clicking buttons requires. This was also the case with classification processes, GEE has a good repository in which users can access and customized allowing users with no programming background to easily keep up with the documentation and coding. GEE performance in time consumption is very dependable to internet connection strength and continuity.

Overall, GEE has succeeded to fulfilled all objective required within the designed scopes and limitation. Although, GEE is a very promising platform to perform large-scale processing and geospatial calculation, users still need to understand the fundamentals behind image processing to ensure the quality and reliability of the analysis performed. A basic but structured methodology is the way to go if users want to optimize GEE platform in order to avoid error and reduction in output quality.

The satellite data used has a spatial resolution of 30m, so it is difficult to detect changes within the allocated studied area example can be seen in the misclassified paddy field although the training and testing samples collected for this study were carefully selected. These misclassified pixels also caused by the limitation of seasonal dataset of the area.

However, for future studies, it is recommended to perform more accurate analysis at micro level with fine resolution data with ground truth details. An object-based classification and a fusion on spectral indices need to be included and should also be considered to avoid improve classification and accuracy. Further analysis of the land surface temperature and closer intervals should be done in the future to study the correlation and change pattern between LUCC with urban heat island.

Acknowledgements
Authors would like to thank Malaysian Space Agency (MYSA), Universiti Putra Malaysia (UPM), PLANMalaysia Selangor, MaCGDI and individual advisors for their support and guidance throughout the project.

References
[1] Un-habitat. (2010). State of the world's cities 2010/2011: bridging the urban divide. Earthscan.
[2] Desa, U. N. (2014). World urbanization prospects, the 2011 revision. Population Division, Department of Economic and Social Affairs, United Nations Secretariat.
[3] Chadchan, J., & Shankar, R. (2012). An analysis of urban growth trends in the post-economic reforms period in India. International Journal of Sustainable Built Environment, 1(1), 36-49.
[4] Parry, J. A., Ganaie, S. A., & Bhat, M. S. (2018). GIS based land suitability analysis using AHP model for urban services planning in Srinagar and Jammu urban centers of J&K, India. Journal of Urban Management.
[5] Güneralp, B., & Seto, K. C. (2008). Environmental impacts of urban growth from an integrated dynamic perspective: A case study of Shenzhen, South China. Global Environmental Change,
[6] Noor, N. M., Rosni, N. A., Hashim, M., & Abdullah, A. (2018). Developing land use geospatial indices (LUGI) for sprawl measurement in alpha cities: Case study of Kuala Lumpur, Malaysia. Cities.

[7] Dewan, A. M., & Yamaguchi, Y. (2009). Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. Applied Geography, 29(3), 390-401.

[8] Chan, K. M., & Vu, T. T. (2017). A landscape ecological perspective of the impacts of urbanization on urban green spaces in the Klang Valley. Applied Geography, 85, 89-100.

[9] Singh, P., Kikon, N., & Verma, P. (2017). Impact of land use change and urbanization on urban heat island in Lucknow city, Central India. A remote sensing based estimate. Sustainable cities and society, 32, 100-114.

[10] Lee, J., & Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. Big Data Research, 2(2), 74-81. doi: 10.1016/j.bdr.2015.01.003

[11] Yang, C., Yu, M., Hu, F., Jiang, Y., & Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. Computers, environment and urban systems, 61, 120-128.

[12] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27.

[13] Pimple, U., Simonetti, D., Sitthi, A., Pungkul, S., Leadprathom, K., & Skupek, H. et al. (2018). Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thailand. Journal Of Computer And Communications, 06(01), 247-264. doi: 10.4236/jcc.2018.61025

[14] Azzari, G., & Lobell, D. (2017). Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. Remote Sensing Of Environment, 202, 64-74. doi: 10.1016/j.rse.2017.05.025

[15] Goldblatt, R., Deininger, K., & Hanson, G. (2018). Utilizing publicly available satellite data for urban research: Mapping built-up land cover and land use in Ho Chi Minh City, Vietnam. Development Engineering, 3, 83-99. doi: 10.1016/j.deveng.2018.03.001

[16] Celic, N. (2018, July). Change Detection of Urban Areas in Ankara through Google Earth Engine. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP) (pp. 1-5). IEEE.

[17] Goldblatt, R., You, W., Hanson, G., & Khandelwal, A. (2016). Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. Remote Sensing, 8(8), 634. doi: 10.3390/rs8080634

[18] Kerajaan Negeri Selangor. (2018). Sejarah negeri Selangor. Portal Kerajaan Negeri Selangor. Retrieved December 2018.

[19] GDP By State - National Accounts - 2010-2015". Department of Statistics, Malaysia. Department of Statistics, Malaysia. p. 10. Retrieved December 2018.

[20] Population by States and Ethnic Group. (2015). Department of Information, Ministry of Communications and Multimedia, Malaysia. Retrieved December 2018.

[21] Wulder M.A., Masek J.G., Cohen W.B., Loveland T.R., Woodcock C.E. (2012): Opening the archive - How free data has enabled the science and monitoring promise of Landsat. Remote Sensing of Environment. Vol 122, pp. 2-10

[22] Anderson, J. R. (1976). A land use and land cover classification system for use with remote sensor data (Vol. 964). US Government Printing Office.

[23] Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user’s perspective. Photogrammetric Engineering and remote sensing, 52(3), 397-399.

[24] Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., ... & Moore III, B. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote sensing of environment, 185, 142-154.
[25] Oliphant, A. J., Thenkabail, P. S., Teluguntla, P., Xiong, J., Gumma, M. K., Congalton, R. G., & Yadav, K. (2019). Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. International Journal of Applied Earth Observation and Geoinformation, 81, 110-124.