Machine Learning-Based Land Cover Change Analysis in Surabaya Using Google Earth Engine

Devika Ayu Setyaningrum[1], Ervin Yohannes[2]

[1,2] (Informatics Engineering, Universitas Negeri Surabaya)
[1]24051204255@mhs.unesa.ac.id
[2] ervinyohannes@unesa.ac.id

***Abstract*— Surabaya City has experienced rapid urban development, resulting in significant land cover changes. Currently, land cover monitoring is still largely conducted manually, creating a need for a more effective, efficient, and spatially based monitoring approach through remote sensing technology. These changes can be efficiently monitored using cloud-based remote sensing technology through the Google Earth Engine (GEE) platform. This platform enables geospatial data analysis by utilizing Google's data repository and implementing algorithms through JavaScript and Python programming languages. This study aims to identify land cover dynamics in Surabaya City between 2016 and 2024 using the Random Forest machine learning algorithm based on Landsat 8 OLI/TIRS imagery within the GEE platform. To improve classification performance, several spectral indices were incorporated, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), and Normalized Difference Water Index (NDWI). Land cover was classified into four categories: water bodies, vegetation, built-up areas, and bare land. The classification results indicate that vegetation was the dominant land cover class in 2016, accounting for 36.28% of the total area. In contrast, built-up areas became the dominant class in 2024, representing 35.62% of the total area. Analysis of land cover changes between 2016 and 2024 revealed a decrease of 2,453.95 hectares in vegetation cover, while built-up areas increased by 1,756.40 hectares. The application of the GEE platform for land cover identification using the Random Forest method proved to be effective and efficient, as demonstrated by an overall classification accuracy exceeding 85%.**

Keyword*— Land Cover Change, Remote Sensing, Google Earth Engine, Random Forest, Landsat-8

I. INTRODUCTION

Surabaya is one of the metropolitan cities in Indonesia that is experiencing very rapid development and construction. As the center of economic, industrial and government activities in East Java Province, Surabaya is experiencing very rapid regional growth dynamics, both in terms of infrastructure development, settlements, and industrial areas. The increase in population in the city area can increase energy consumption used for household, transportation and industrial purposes. So that land use will shift from agricultural purposes to residential purposes and so on. This change is generally characterized by an increase in built-up areas, a reduction in green open spaces, and changes in land use that have an impact on environmental conditions and the sustainability of the city [13].

The continous land conversion driven by urban growth is difficult to monitor using conventional manual methods. Therefore, effective and efficient monitoring approaches based on spatial technologies are required to identify land cover dynamics[1]. Remote sensing technology utilizing satellite imagery has become a suitable solution because it provides multi-temporal spatial data with broad spatial coverage and appropriate spatial resolution [5]. In recent years, remote sensing technology has increasingly adopted cloud computing platforms, one of which is Google Earth Engine (GEE), introduced by Google in 2010 [17].

Google Earth Engine (GEE) is a cloud-based platform designed for processing and analyzing geospatial data by utilizing Google's extensive geospatial data repository [20]. GEE supports JavaScript and Python programming languages for implementing various geospatial analysis algorithms [4]. The platform is equipped with algorithms specifically developed to process georeferenced satellite imagery stored in the cloud, enabling efficient utilization of satellite data for a wide range of spatial analyses while simplifying data management. Furthermore, GEE provides access to a comprehensive global satellite data catalog, including Landsat-8 imagery [9]. Through its cloud-based data management capabilities, GEE facilitates land cover identification and classification for monitoring and mapping purposes [24]. Moreover, the integration of machine learning algorithms such as Random Forest (RF) within GEE enables accurate land cover classification and subsequent analysis of land cover changes over time [22].

Numerous studies have investigated the application of the Google Earth Engine platform for land cover classification using the Random Forest machine learning algorithm. These studies have demonstrated that the integration of GEE and RF offers high classification accuracy and computational efficiency for satellite image analysis. Fathoni H.A. (2025) [9] conducted land cover mapping in the Nusantara Capital City (IKN) area using the Random Forest algorithm. The study classified land cover into five categories: water bodies, built-up areas, bare land, natural vegetation, and cultivated vegetation, achieving an overall accuracy of 81.29%. Similarly, Febriani et al. (2022) [10] analyzed land cover dynamics using Landsat imagery on the Google Earth Engine platform with the Random Forest algorithm for the years 2013, 2017, and 2021.

Based on these two studies, it can be seen that the Random Forest algorithm is considered quite effective in handling high-dimensional data, resistant to overfitting and capable of processing multispectral data in parallel in cloud computing environments such as Google Earth Engine (GEE) [3]. Therefore, this study proposes the use of the Google Earth Engine (GEE) cloud computing platform combined with the

Random Forest (RF) machine learning algorithm in analyzing the dynamics of land cover changes in Surabaya City in 2016 and 2024. In this study, the spectral index on Landsat-8 satellite imagery is utilized to strengthen the classification process, the spectral indices used include, Normalize Difference Vegetation Index (NDVI), and Normalize Difference Built-Up Index (NDBI) and Normalize Difference Water Index (NDWI). This study is expected to provide a more detailed picture of the spatial distribution pattern and temporal changes in land cover in Surabaya during 2016 and 2024. The results of this analysis are expected to be a scientific basis for local governments in adaptive urban planning to climate change, for example through the provision of green open spaces (RTH) and control of building density.

II. METHODS

This study employed a quantitative research approach, as all data were analyzed in numerical form using satellite imagery (Landsat 8). The data consisted of spectral reflectance values, the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), and the Normalized Difference Water Index (NDWI), which were used as input variables for the machine learning-based land cover classification process.

This study is categorized as spatial and multi-temporal research, as it focuses on comparing land cover conditions at two different time periods, namely 2016 and 2024, to identify land cover dynamics and spatial transformation in Surabaya City. The study employs a spatio-temporal analysis approach based on remote sensing by implementing the Random Forest machine learning algorithm on the Google Earth Engine (GEE) platform to analyze land cover dynamics. This study is expected to produce a land cover distribution map, provide information on the temporal trends of land cover change between 2016 and 2024, and identify areas that have experienced significant land cover changes within Surabaya City.

A. Research Site

The study was conducted in Surabaya City, East Java Province, Indonesia. Surabaya the capital city of East Java Province, is geographically located between 7°09′–7°21′ South Latitude and 112°36′–112°54′ East Longitude. According to the Decree of the Minister of Home Affairs of the Republic of Indonesia No. 100.1.1-117 of 2022 concerning the Assignment and Updating of Administrative Region and Island Codes and Data, the total area of Surabaya City is 335.925 km². Surabaya is bounded by the Java Sea and the Madura Strait to the north, Sidoarjo Regency to the south, Gresik Regency to the west, and the Madura Strait to the east.

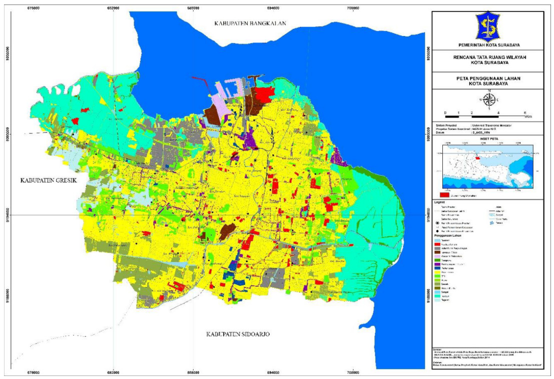


Fig. 1 Land Use Map of Surabaya City [21]

B. Data Used

The data used in this study included Landsat 8 Operational Land Imager/ Thermal Infrared Sensor (OLI/TIRS) Collection 2 Level 2 Satellite Imagery for the years 2016 and 2024, which was accessed and processed through the Google Earth Engine (GEE) platform. A vector shapefile of the administrative boundary of Surabaya City was used to define the study area and delimit the spatial extent of the analysis. In addition, a shapefile containing delineated land cover sample areas was used as reference data for validating the land cover classification model. This investigation was conducted using Asus Laptops (8GB of RAM), Google Maps, and Microsoft Excel.

C. Research Flowchart

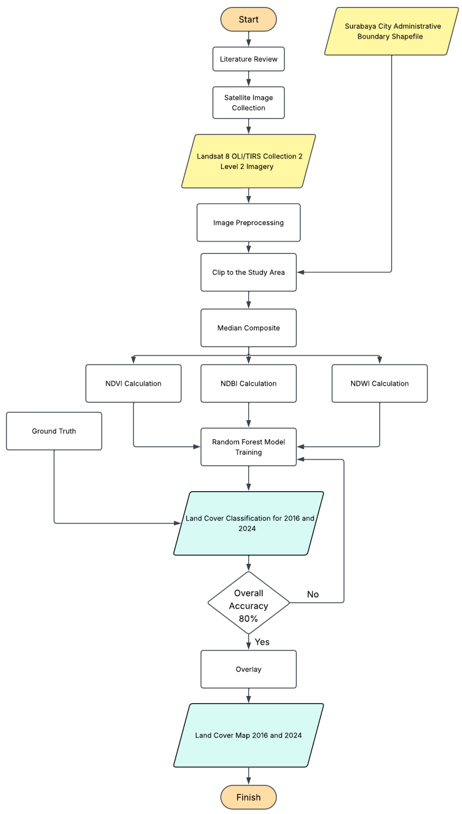


Fig.2 Schematic diagram

Based on Fig. 2, the research workflow consists of several stages, including the acquisition of multi-temporal satellite imagery, image preprocessing (radiometric correction and cloud masking), calculation of spectral indices such as the Normalized Difference Vegetation Index (NDVI), Normalize Difference Built-Up Index and Normalize Difference Water Index as supporting variables, preparation of training and testing datasets, training of the machine learning model, accuracy assessment using a confusion matrix, and the generation of land cover classification maps followed by change detection analysis to evaluate land cover changes over time.

D. Data Preprocessing

In this study, the image preprocessing stage was carried out using the Google Earth Engine (GEE) platform. The preprocessing workflow consisted of several main steps, including defining the study area, acquiring satellite imagery, applying cloud masking and median compositing, and applying the scale. The administrative boundary of Surabaya City was used as the Region of Interest (ROI), which served as the analysis area for the image processing workflow. In the Google Earth Engine (GEE) platform, the administrative boundary shapefile of Surabaya City was uploaded to the Legacy Assets repository and used as the basis for clipping the satellite imagery to the study area.

Subsequently, Landsat 8 OLI/TIRS Collection 2 Level 2 satellite imagery for the years 2016 and 2024 was acquired from the Google Earth Engine (GEE) data catalog. The imagery used in this study covered the period from 1 January to 31 December for each observation year, namely 2016 and 2024. Following image acquisition, cloud masking was performed using the Quality Assessment (QA_PIXEL) band of the Landsat 8 imagery to remove pixels identified as clouds, cloud shadows, and other sources of noise. This process aimed to produce clean imagery that more accurately represented the actual surface conditions. The cloud-masked images were subsequently processed using median compositing, in which the median pixel value was calculated from all available images within each observation year to generate a single representative composite image.

The final preprocessing step involved the application of the scale factor. Landsat Collection 2 Level 2 products are stored as scaled integer values to improve storage efficiency. Therefore, the scale factor and offset provided with each spectral band were applied to convert the stored digital values back to their original floating-point reflectance values before further analysis. This process ensured that the spectral information used in subsequent calculations accurately represented the surface reflectance recorded by the sensor.

E. Data Processing

The data processing stage followed the image preprocessing stage and aimed to extract spectral information relevant to land cover characteristics. In this study, the data processing stage involved the calculation of several spectral indices, including the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), and the Normalized Difference Water Index (NDWI).

1) *NDVI (Normalize Difference Vegetation Index)*

The NDVI was calculated to assess vegetation density based on the reflectance values of the Near-Infrared (NIR) and Red bands. The calculaton of NDVI is performed using the following equation 1 :

$$NDVI = \frac{(\rho NIR - \rho RED)}{(\rho NIR + \rho RED)} \tag{1}$$

where :
ρNIR = the reflectance value of Near-Infrared band

ρRED = the reflectance value of red band

The classification range of NDVI values range from -1 to 1. Objects with NDVI values ranging from -1 to 0 are classified as non-vegetated surfaces, whereas values approaching +1 indicate surfaces with high vegetation density. [38]

2) *NDBI (Normalize Difference Built-Up Index)*

The Normalized Difference Built-up Index (NDBI) is a spectral index used to identify and assess the density of built-up areas based on the reflectance values of the Shortwave Infrared (SWIR) and Near-Infrared (NIR) bands. The calculation of NDBI is performed using the following equation 2:

$$NDBI = \frac{(\rho SWIR - \rho NIR)}{(\rho SWIR + \rho NIR)} \tag{2}$$

where :
ρSWIR = the reflectance value of Shortwave infrared

ρNIR = the reflectance value of Near Infrared

The NDBI values range from -1 to +1, where negative values generally correspond to water bodies, whereas higher positive values indicate a greater extent of built-up areas. [29]

3) *NDWI (Normalize Difference Water Index)*

The Normalized Difference Water Index (NDWI) is a spectral index used to detect the presence of water bodies and monitor surface water conditions. NDWI is calculated using the reflectance values of the Green band (Band 3) and the Near-Infrared (NIR) band (Band 5) of Landsat 8 imagery [29]. The calculation of NDWI is performed using following equation 3.

$$NDWI = \frac{(\rho Green - \rho NIR)}{(\rho Green + \rho NIR)} \tag{3}$$

where :
ρGreen = the reflectance value of green band

ρNIR = the reflectance value of Near Infrared

After the NDVI, NDBI, and NDWI spectral indices were calculated, they were integrated with the original spectral bands of the Landsat 8 imagery to produce a *feature stack*. This feature stack represents a multiband image containing all spectral variables and derived indices, which served as the input features for the machine learning-based land cover classification process.

In this study, land cover was classified into four categories: built-up areas, vegetation, water bodies, and bare land. The subsequent step involved generating the land cover class labels (*LULC_Class*). The class labels were assigned based on predefined threshold values for each spectral index. This process produced labeled data, which were subsequently used as the target class labels for training the Random Forest classification algorithm.

In the land cover classification process using the Random Forest algorithm, the NDVI, NDBI, and NDWI indices were used simultaneously as input features to enable the model to distinguish the characteristics of different land cover classes. The implementation of the Random Forest machine learning model began with pixel sampling, which was used to generate the training and testing datasets. In this study, a total of 5,000 pixels were randomly sampled at a spatial resolution of 30 m, corresponding to the spatial resolution of the Landsat 8 imagery used. A seed value of 42 was specified during the sampling process to ensure reproducibility and to generate a consistent distribution of sample pixels. The sampled data were then divided into training and testing datasets, with 70% (3,500 pixels) allocated for model training and the remaining 30% (1,500 pixels) reserved for model testing.

The Random Forest classifier was constructed using 100 decision trees through the function in Google Earth Engine *ee.Classifier.smileRandomForest(100).* The model was trained using the training dataset, with NDVI, NDBI, and NDWI serving as the predictor variables (input features), while *LULC_Class* was used as the target variable (class label). Each decision tree learned the relationship between the combination of these three spectral indices and the corresponding land cover classes.

F. Data Analysis and Evaluation

The data analysis and evaluation stage was conducted after the land cover classification process using the Random Forest algorithm. At this stage, the classification results were analyzed to assess the model's ability to accurately identify each land cover class. Model evaluation is essential to ensure that the resulting land cover map has a high level of reliability and can be used for subsequent spatial and temporal land cover change analysis.

The accuracy assessment was performed using ground truth data obtained through visual interpretation of high-resolution satellite imagery and supporting reference maps. These validation data were then compared with the classification results to construct a confusion matrix. The confusion matrix was used to evaluate the accuracy of the land cover classification by measuring the agreement between the predicted classes and the corresponding reference classes. The evaluation metrics derived from the confusion matrix included Producer's Accuracy (PA), User's Accuracy (UA), Overall Accuracy (OA), and the Kappa coefficient. In this study, the classification results were considered satisfactory when the overall accuracy exceeded 80%.

Based on the confusion matrix, the Overall Accuracy (OA) and Kappa coefficient were calculated as indicators of the classification model's performance. Overall Accuracy represents the percentage of correctly classified pixels relative to the total number of validation samples. Meanwhile, the Kappa coefficient measures the level of agreement between the classification results and the reference data while accounting for agreement that may occur by chance. The Kappa coefficient ranges from 0 to 1, with values closer to 1 indicating higher classification accuracy and stronger agreement between the predicted and reference classes.

TABLE I
CONFUSSION MATRIX

Data Refrerence	Klasifikasi				Jumlah	*Producer's Accuracy*
	1	2	3	4		
1	x_{11}	x_{12}	x_{13}	x_{14}	x_{1+}	x_{11}/ x_{1+}
2	x_{21}	x_{22}	x_{23}	x_{24}	x_{2+}	x_{22}/ x_{2+}
3	x_{31}	x_{32}	x_{33}	x_{34}	x_{3+}	x_{33}/ x_{3+}
4	x_{41}	x_{42}	x_{43}	x_{44}	x_{4+}	x_{44}/ x_{4+}
Jumlah	x_{1+}	x_{2+}	x_{3+}	x_{4+}	n	
User's Accuracy	x_{11}/ x_{1+}	x_{22}/ x_{2+}	x_{33}/ x_{3+}	x_{44}/ x_{4+}		

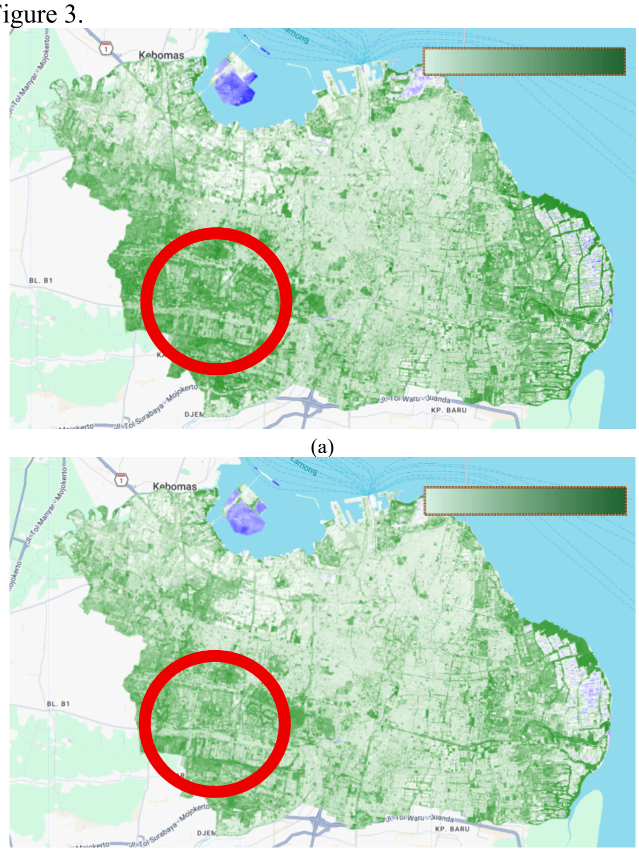
In this study, the classification model was considered to have satisfactory performance when the Overall Accuracy (OA) reached at least 80% and was supported by a high Kappa coefficient. If the classification accuracy did not meet this threshold, the model was optimized by adjusting the Random Forest parameters, increasing the number of training samples, or selecting more representative input variables, including combinations of spectral bands and spectral indices such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), and Normalized Difference Water Index (NDWI).

III. RESULTS AND DISCUSSION

Land cover classification in this study was performed using Landsat 8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) satellite imagery obtained through the Google Earth Engine (GEE) platform. The analysis focused on two observation years, namely 2016 and 2024, to identify the spatial and temporal dynamics of land cover change. Each spectral band of the Landsat 8 imagery records different surface reflectance values for each geographic location, providing the spectral information required to distinguish between different land cover types. In this study, the land cover classes were categorized into four classes: water bodies (Class 0), vegetation (Class 1), built-up areas (Class 2), and bare land (Class 3) within Surabaya City.

A. Results of Vegetation Index (NDVI)

The Google Earth Engine (GEE) platform uses the JavaScript programming language through the Google Earth Engine Application Programming Interface (API) to perform satellite image processing and spatial analysis. In this study, the NDVI algorithm was implemented in Google Earth Engine using the normalizedDifference() function. The vegetation density extraction results for 2016 and 2024 are presented in Figure 3.



(a)

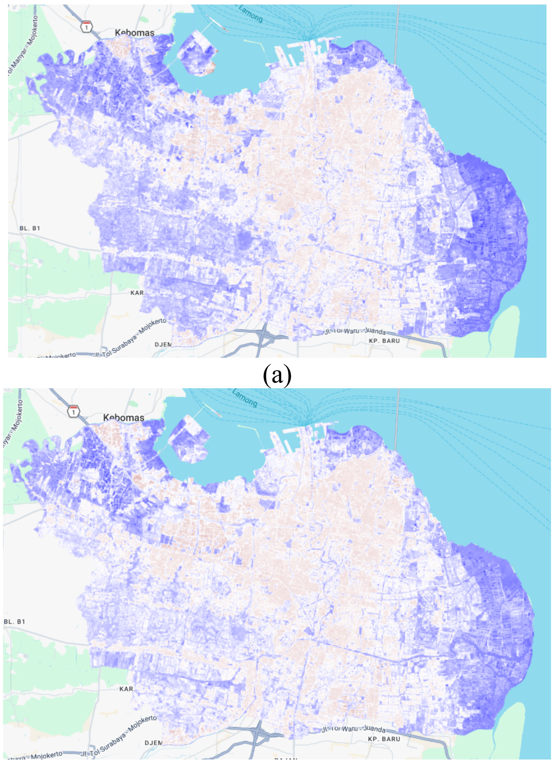
(b)

Fig.3 Result of NDVI (a) 2016 and (b) 2024

As shown in Figure 3, darker shades of green represent areas with higher vegetation density, whereas lighter shades of green approaching white indicate lower vegetation density. In addition, the blue color represents water bodies. A visual comparison of the NDVI maps indicates a noticeable change in vegetation density, particularly in the western part of Surabaya. In the 2024 map, the green tones appear lighter than those in 2016, suggesting a decline in vegetation density in this area. Figure 3 also shows blue-colored areas, which represent water bodies. A visual comparison of the NDVI maps reveals a noticeable change in vegetation density, particularly in the western part of Surabaya. In the 2024 map, the green shades appear lighter than those in 2016, indicating a decline in vegetation density in this area.

B. Results of Built-Up Index (*NDBI*)

In this study, the NDVI was combined with other spectral indices, particularly the Normalized Difference Built-up Index (NDBI), to improve the ability of the machine learning model to accurately recognize and distinguish different land cover patterns. In remote sensing, the Normalized Difference Built-up Index (NDBI) is widely used to identify built-up areas, such as residential areas, buildings, roads, and other urban infrastructure, based on their spectral characteristics. The built-up area index results for 2016 and 2024 are presented in Figure 4.



(a)

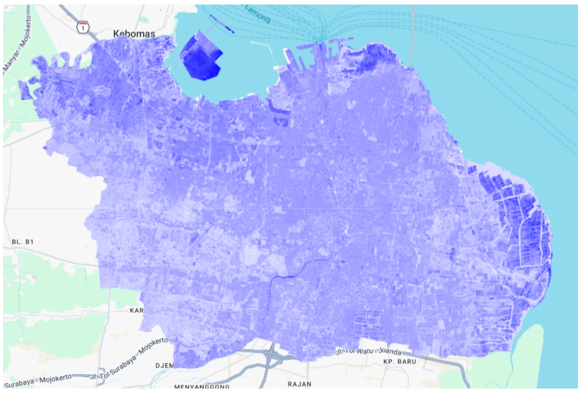
(b)

Fig.4 Result of NDBI (a) 2016 and (b) 2024

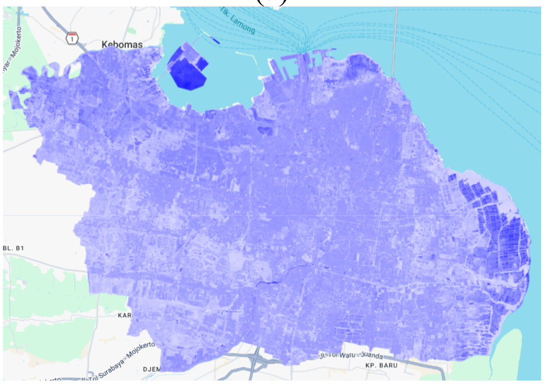
As shown in Figure 4, brighter (whiter) areas indicate a higher density of built-up land, whereas areas with colors approaching blue represent lower built-up density. Spectrally, the NDBI exhibits an inverse relationship with the NDVI, where areas with high NDBI values generally correspond to low NDVI values.

C. Results of Water Index (*NDWI*)

In this study, water bodies were identified using the Normalized Difference Water Index (NDWI) algorithm. NDWI is commonly used to detect the presence of water bodies, such as rivers, reservoirs, fish ponds, and other surface water features, based on their spectral characteristics. The algorithm can also be used to assess surface water conditions. The NDWI implementation used in this study on the Google Earth Engine (GEE) platform is presented below.



(a)



(b)

Fig. 5 Result of NDWI (a) 2016 and (b) 2024

Based on Figure 5, visual interpretation indicates no significant differences between the NDWI results for 2016 and 2024. This suggests that land cover changes associated with water bodies were relatively limited over the study period.

D. Land Cover Classification Results using Random Forest

The predictions generated by all decision trees were aggregated using a majority voting mechanism, in which the class receiving the highest number of votes was assigned as the final classification result for each pixel. The resulting land cover classification maps for 2016 and 2024 are presented in Figure 6

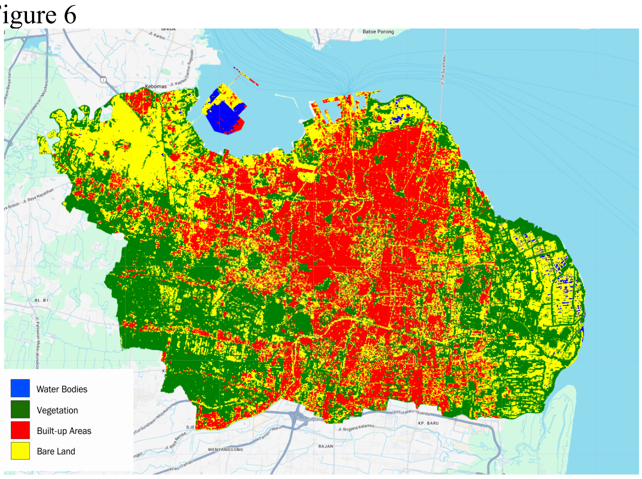


Fig. 6 Land Cover Classification for 2016

Based on the land cover classification results shown in Figure 4.14, vegetation was the dominant land cover class across Surabaya City in 2016, particularly in the western part of the city, where vegetation density remained relatively high. Bare land also occupied a considerable area, especially in eastern Surabaya and the northern part of western Surabaya, including Pakal and Benowo Districts, where large areas were still used as fish ponds and open land. In contrast, built-up areas were primarily concentrated in northern and central Surabaya, which serve as the city's main administrative, commercial, and urban infrastructure centers.

In 2016, the land cover of Surabaya City was dominated by vegetation, covering 12,297.047 ha (36.28%), followed by bare land with an area of 10,866.844 ha (32.06%). Built-up areas occupied 10,315.794 ha (30.44%), while water bodies accounted for the smallest proportion, covering 405.956 ha (1.19%).

TABLE III
AREA OF LAND COVER CLASSES IN SURABAYA CITY IN 2016

Class	Area (Ha)	Percentage (%)
Water bodies	405.956	1.19
Vegetation	12297.047	36.28
Built-Up Area	10315.794	30.44
Bare Land	10866.844	32.06

To identify the dynamics of land cover change, the 2024 land cover classification results are presented in Figure 6 for comparison with the land cover conditions observed in 2016.

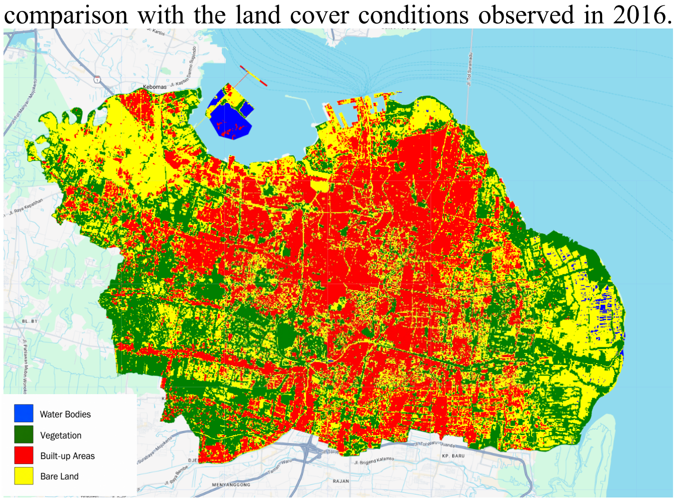


Fig. 7 Land Cover Classification for 2016

The 2024 land cover classification map reveals substantial changes in land cover dynamics compared with 2016. Built-up areas expanded considerably toward the western part of Surabaya, as indicated by the increased dominance of the built-up land cover class in this region. Conversely, the extent of vegetation decreased due to the conversion of vegetated land into built-up areas and bare land.

TABLE IIIII
AREA OF LAND COVER CLASSES IN SURABAYA CITY IN 2024

Class	Area (Ha)	Percentage (%)
Water bodies	371.351	1.09
Vegetation	9844.091	29.04
Built-Up Area	12072.198	35.62
Bare Land	11600.343	34.23

In 2024, the land cover of Surabaya City was dominated by built-up areas, covering 12,072.198 ha (35.62%), followed by bare land with an area of 11,600.343 ha (34.23%). Vegetation covered 9,844.091 ha (29.04%), while water bodies occupied 371.351 ha (1.09%), primarily identified in reclaimed coastal areas.

E. Land Cover Change Analysis

Based on Figures 6 and 7, Surabaya City was predominantly covered by vegetation in 2016, whereas built-up areas became the dominant land cover class in 2024. The changes in land cover between 2016 and 2024 for each land cover class are presented below.

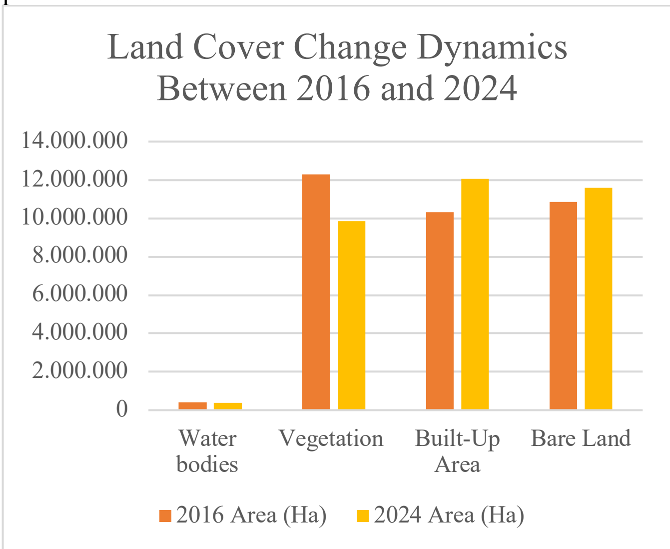


Fig. 8 Land Cover Change Graph

Based on Figure 8, the vegetation land cover class experienced a substantial change. By 2024, the area covered by vegetation had decreased by approximately 2,453.95 ha. In contrast, built-up areas increased by 1,756.40 ha, indicating significant urban expansion in Surabaya City. The continuous increase in urban development has contributed to the reduction of vegetated areas, with much of the land being converted into built-up areas and bare land. Figure 4.3 also illustrates the land cover changes between 2016 and 2024, which are further presented using a gain-and-loss analysis chart.

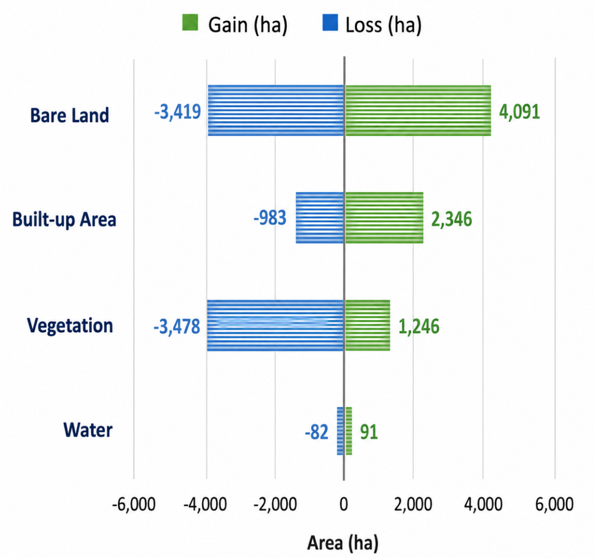


Fig. 9 Gain and Loss of Land Cover Classes Between 2016 and 2024

Figure 9 presents the land cover change analysis in hectares. The green bars represent gains, whereas the purple bars indicate losses. Based on Figure 9, the largest increase occurred in the bare land class. This increase may be attributed to the development of new areas, temporary land-use conversion, or the expansion of underutilized land.

A substantial increase was also observed in the built-up area class, with a gain of 2,641.90 ha. This finding indicates rapid expansion of residential areas, industrial zones, and urban infrastructure. In contrast, the vegetation class experienced the greatest decline, with a loss of 3,753.21 ha, while the gain was only approximately 1,300 ha. These results suggest that a considerable proportion of vegetated land has been converted into built-up areas and bare land. If this trend continues, it may lead to environmental degradation, including increased land surface temperatures, reduced water infiltration capacity, and a higher risk of flooding.

F. Classification Accuracy Result

Based on the implementation of the Random Forest algorithm on Landsat 8 satellite imagery processed using the Google Earth Engine (GEE) platform, the land cover map of Surabaya City was successfully classified into four major classes: water bodies, vegetation, bare land, and built-up areas. The accuracy assessment of the 2016 and 2024 land cover maps was conducted using a confusion matrix generated by comparing the classified pixel values with reference sample points.

The confusion matrix was used to calculate several accuracy metrics, including Overall Accuracy (OA), Producer's Accuracy (PA), User's Accuracy (UA), and the Kappa coefficient. These metrics were used to evaluate the performance and reliability of the land cover classification results. In this study, a total of 40 validation samples were used, with each land cover class represented by 10 randomly selected sample points. Based on the accuracy assessment, the land cover classification accuracy matrices are presented in Tables IV

TABLE IVV
CONFUSION MATRIX FOR THE 2024 LAND COVER CLAIFICATION

Classification Class	Ground Truth				Row Total	UA (%)
	Water bodies	Vegetation	Built-Up Area	Bare Land		
Water bodies	8	0	0	2	10	80
Vegetation	0	9	0	1	10	90
Built-Up Area	0	0	10	0	10	100
Bare Land	1	0	1	8	10	80
Column total	9	9	11	10	40	
PA (%)	88.88	100	90.9090909	80		
OA (%)	87.5					
Kappa	84.49612403					

TABLE V
CONFUSION MATRIX FOR THE 2024 LAND COVER CLAIFICATION

Classification Class	Ground Truth				Row Total	UA (%)
	Water bodies	Vegetation	Built-Up Area	Bare Land		
Water bodies	8	0	0	2	10	80
Vegetation	0	9	0	1	10	90
Built-Up Area	0	0	9	1	10	90
Bare Land	1	0	1	8	10	80
Column total	9	9	11	10	40	
PA (%)	88.888	100	81.8181818	80		
OA (%)	85					
Kappa	81.39534884					

Based on Tables IV and table V, the overall accuracy of the 2024 land cover classification reached 87%, with a Kappa coefficient of 84.49%. In comparison, the 2016 land cover classification achieved an overall accuracy of 85% and a Kappa coefficient of 81.39%. The Producer's Accuracy (PA) and User's Accuracy (UA) values varied among the different land cover classes. In both 2016 and 2024, the highest Producer's Accuracy was obtained for the vegetation class, whereas the lowest value was recorded for the bare land class at 80%.

A different pattern was observed for User's Accuracy. In 2024, the highest User's Accuracy was achieved by the built-up area class, whereas in 2016, both the vegetation and built-up area classes exhibited the highest User's Accuracy values. As shown in Tables IV and V, the relatively low Producer's Accuracy and User's Accuracy for the bare land class were associated with the confusion matrix results, which indicated that a number of bare land pixels were misclassified as water bodies and built-up areas.

These classification errors were likely caused by the similarity in spectral characteristics among different land cover classes in Landsat 8 imagery, which has a spatial resolution of 30 m. In addition, a single pixel may represent more than one land cover type, such as a mixture of bare land, built-up areas, and vegetation. Consequently, some pixels exhibit similar spectral signatures, making them more difficult to classify accurately.

Overall, the accuracy assessment demonstrates that the Random Forest classification model achieved good classification performance and is sufficiently reliable for land cover mapping and subsequent land cover change analysis.

IV. CONCLUSION

Based on the results of the analysis and discussion of the research conducted, the following conclusions can be drawn:

1. The identification and mapping of land cover in the City of Surabaya for 2016 and 2024 can be carried out effectively using Landsat-8 satellite imagery processed on the Google Earth Engine platform. The classification results successfully identified four land cover classes: vegetation, bare land, built-up area, and water body, which were then visualized in the form of land cover maps.
2. The analysis of land cover change indicates land use dynamics during the 2016–2024 period. The most prominent changes are the increase in built-up areas and bare land, along with a decrease in vegetation area. Meanwhile, changes in the water body class were relatively small. This condition indicates urban development that has led to a reduction in vegetated areas.
3. The implementation of machine learning algorithms on Google Earth Engine enables automatic and efficient land cover classification based on the spectral characteristics of Landsat-8 imagery. The algorithms used were able to distinguish each land cover class and produced classification maps that can be utilized for spatial analysis of land cover change.
4. The accuracy evaluation using a confusion matrix shows that the classification results have a good level of accuracy. For 2016, the classification achieved an overall accuracy of 85% and a Kappa coefficient of 81.39%. For 2024, it achieved an overall accuracy of 87.5% and a Kappa coefficient of 84.49%. These results indicate that the classification model used is sufficiently reliable to support land cover change analysis in the City of Surabaya.

REFERENCE

[1] Akbar, R. (2025). Identifikasi perubahan tutupan lahan Kabupaten Kendal menggunakan Google Earth Engine. *Indonesian Journal of Spatial Planning*, 6(2), 43–49.

[2] Fathoni, H. A. (2025). *Pemetaan tutupan lahan menggunakan Random Forest pada platform Google*

Earth Engine: Studi kasus IKN. Prosiding TECHNOPEX-2025, Institut Teknologi Indonesia. ISSN: 2654-489X.

[3] Febriani, N., Yunidar, S., Hidayat, R. A., Amor, G., & Indrayani, P. (2022). Klasifikasi citra satelit dengan metode random forest untuk observasi dinamika lanskap ekosistem Kabupaten Sijunjung. *El-Jughrafiyah*, 2(2).

[4] Fikri, A. S., Setiawan, F., Violando, W. A., Muttaqin, A. D., & Rahmawan, F. (2021). Analisis Penutupan Lahan menggunakan Google Earth Engine (GEE) dengan Metode Klasifikasi Terbimbing (Studi kasus: WIilayah Pesisir Kabupaten Lamongan, Provinsi Jawa Timur). Prosiding FIT ISI, 1, 89–99. http://eprints.itn.ac.id/1418/

[5] Firmawan MA, Widiatmaka, Nirmala K. 2021. Identifikasi dinamika spasial penggunaan dan tutupan lahan di Kabupaten Indramayu. Jurnal Ilmu Tanah dan Lingkungan. 23(2): https://doi.org/10.29244/jitl.23.2.78-84

[6] Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google earth engine: Planetary-scale geospatial analysis for everyone,” Remote Sens. Environ., vol. 202, pp. 18–27, Dec. 2017.

[7] Hansen et al., “High-resolution global maps of 21st-century forest cover change,” Science, vol. 342, no. 6160, pp. 850–853, Nov. 2013.

[8] Kumar and O. Mutanga, “Google Earth Engine applications since inception: Usage, trends, and potential,” Remote Sens., vol. 10, no. 10, Sep. 2018, Art. no. 1509.

[9] Lillesand, Thomas, M., 1990, Penginderaan Jauh dan Interpretasi Citra, Penerj. Stanto, Gadjah Mada University Press : Yogyakarta.

[10] Noviani, T. C., Armijon, Tridawati, A., & Samri, A. S. (2024). Analisis perubahan tutupan lahan tahun 2013–2022 di Kota Semarang menggunakan Google Earth Engine. *Jurnal Ilmiah Tekno Global, 13*(1), July 2024.

[11] Pemerintah Kota Surabaya. (2024). *Peraturan Wali Kota Surabaya Nomor 52 Tahun 2024 tentang Rencana Kerja Pemerintah Daerah Kota Surabaya Tahun 2025*. Surabaya: Pemerintah Kota Surabaya. Diakses dari https://surabaya.go.id

[12] Prasetyo, S. T., Rahman, F. A., Suryawati, S., Supriyadi, S., & Setiawan, E. (2025). Land use analysis using machine-learning based on cloud computing platform. *Jurnal Ilmu Pertanian Indonesia (JIPI)*, 30(4), 765–772. https://doi.org/10.18343/jipi.30.4.765

[13] Pratiwi, A. Y., & Jaelani, L. M. (2020). Analisis perubahan distribusi urban heat island (UHI) di Kota Surabaya menggunakan citra satelit Landsat multitemporal. *Jurnal Teknik ITS*, 9(2).

[14] Putri, R. A., et al. (2023). Analisis tutupan lahan menggunakan Google Earth Engine dan citra Landsat 8 OLI (Studi kasus Kabupaten Belitung Timur).

[15] Suryono, H., Marsuhandi, A. H., & Pramana, S. (n.d.). *Klasifikasi tutupan lahan berdasarkan random forest algorithm menggunakan cloud computing platform (Land cover classification based on random forest algorithm using cloud computing platform)*. Badan Pusat Statistik.

[16] Pindarwati, A., Wijayanto, A. W., Rosyani, P., & Maghfiroh, M. F. N. (2025). Integrating satellite imagery and multicriteria decision analysis for high-resolution flood vulnerability mapping: A case study of Jakarta, Indonesia. *International Journal of Advances in Data and Information Systems*, 6(2), 308–319. https://doi.org/10.59395/ijadis.v6i2.1388

[17] Sukoco, B., Armijon, A., & Fadly, R. (2022). Kajian pemanfaatan teknologi Google Earth Engine untuk bidang penginderaan jauh. *Jurnal Penelitian Geografi*, 10(2), 142–149.

[18] U.S. Geological Survey. (2010). Thousands of Landsat scenes in Google’s Earth Engine.https://www.usgs.gov/core-science-systems/nli/landsat/december-5-2010-thousands-landsat-scenes-googles-earth-engine

[19] Wadud, M. V. A., Hermawan, E., & Kamilah, N. (2023). *Analisis Pola Distribusi Spasial Perubahan Penggunaan Lahan dan Urban Heat Island Menggunakan Google Earth Engine (Studi Kasus di Kota Bogor Tahun 2000, 2009, & 2021)*. INFOTECH Journal, Vol. 9 No. 1, 259–269. https://doi.org/10.31949/infotech.v9i1.5507

[20] Fawzi, N.I., 2016, “Koreksi Radiometrik Landsat-8”, *Thermal Remote Sensing Research Center*