Identification of Factors Causing Land Cover Change in the Cikapundung Watershed

Agung Budi Harto[a,b], Riantini Virtriana[a,b], Ricky Jaya Kusuma[b], Ifandra Kusuma Cahya Reynaldi[b], Aufa Qoulan Karima[b], Arno Adi Kuntoro[c]

[a] *Geospatial Information Science and Technology Research Group, Faculty of Earth Science and Technology, Bandung Institute of Technology, Bandung, Indonesia, 40132; virtriana@itb.ac.id (R.V.);*
[b] *Geodesy and Geomatic Engineering, Faculty of Earth Sciences and Technology, Bandung, Indonesia, 40132.; rickyjaya48@gmail.com (R.J.K.); ifandrakusumacahyareynaldi@gmail.com (I.K.C.R.); aufaqoulankarima@gmail.com (A.Q.K);*
[c] *Water Resources Engineering and Management, Faculty of Civil and Environmental Engineering, Bandung, Indonesia, 40132; arnoak@itb.ac.id (A.A.K);*

KEYWORDS: land cover, random forest, land cover change, regression analysis

ABSTRACT:

Land cover change (LCC) is one of the most important global environmental issues in this modern era. Land cover change that occurs at local scales can impact global ecosystems due to cascading effects. The relationship between human population growth, urbanization, and land resource pressures at various spatial and temporal scales has led to these changes. In Bandung City, the increase in population and urbanization over the years has led to a decrease in land resources and a worsening of disasters and climate change. This includes the Cikapundung watershed, where the transformations of green land into built-up areas have caused flooding that affected more than 600 houses in early 2024. This makes it even more important to not only understand historical and current land cover change patterns, but also understand the causes of land cover change and predict future spatiotemporal trends for strategic planning of human settlements, land use, and resource conservation in Bandung City. This study aims to identify the causal factors of long-term land cover change over 50 years (1973-2023) in the Cikapundung watershed and its surroundings. Land cover classification was conducted using random forest algorithm based on Landsat 5, Landsat 8, and Sentinel-2 satellite images. The causal factors of LCC were identified to see the change from non-urban to urban using regression analysis. This study is intended to provide insight into the factors influencing LCC in the Cikapundung watershed and its surroundings so that it can be used to develop effective policies to mitigate the significant impacts of land cover change.

1. INTRODUCTION

Land is essential in sustaining humans and influencing interactions with the natural environment. Changes in land cover, especially from undeveloped to developed land, can disrupt ecosystem patterns and functions, such as climate change, both globally and regionally (Shang et al., 2023). High human activities and needs are factors of land cover change that cause low water retention capacity in the soil and increased surface water runoff (Kuntoro et al., 2017). Thus, this can harm the environment as it has the potential to cause flooding in the watershed area. In the Upper Citarum Watershed, which covers and passes through the Bandung Metropolitan Area, rapid land cover change has occurred due to rapid urbanization, agricultural intensification, forest conversion, and intensive development. This massive change has degraded the watershed, reduced the water quality in the river, and has become one of the factors that trigger flooding in the southern Bandung area, such as Dayeuhkolot, Bojongsoang, and Majalaya (Fardani, 2020).

Land cover change is a phenomenon that takes a long time and involves many natural and social interactions (Liu et al., 2023). In their article, Agaton, et al (2016) mentioned that the changes that have occurred on the earth's surface in the last two decades are the highest due to high human activity. Therefore, it is important to discuss land cover change over a long period to read the pattern of change, as well as to reduce the negative impacts of undue land cover change. In line with this, this study discusses land cover classification modeling in 1973, 1999, and 2023 in the Cikapundung watershed using Landsat images with random forest machine learning and land cover prediction. In addition, this study aims to identify land cover changes, especially from undeveloped land to developed land, and identify factors that cause land cover changes.

2. DATA AND METHOD

2.1 Study Area

The Cikapundung watershed is one of the sub-watersheds of the Upper Citarum watershed, which is the study area of this research. This sub-watershed flows through the Cikapundung River, which has its headwaters at Mount Bukit Tunggul, and follows the valley at Mount Palasari until it turns to the area south of Curug Omas (Sunandar, 2024). The Upper Cikapundung River unites with two other rivers, the Cigulung River and the Cikawari River, dividing Bandung City until it reaches the Citarum River in the south. The Cikapundung River flowing through Bandung has an important role in shaping the city's image and water resources, and serves as a recreation area, drinking water source, and main drainage channel (Yustiani & Lidya, 2016). On the other hand, the river also contributes to flooding in southern Bandung, especially in the low-elevation Upper Citarum watershed areas, such as Dayeuh Kolot and Majalaya (Kuntoro et.al., 2017).

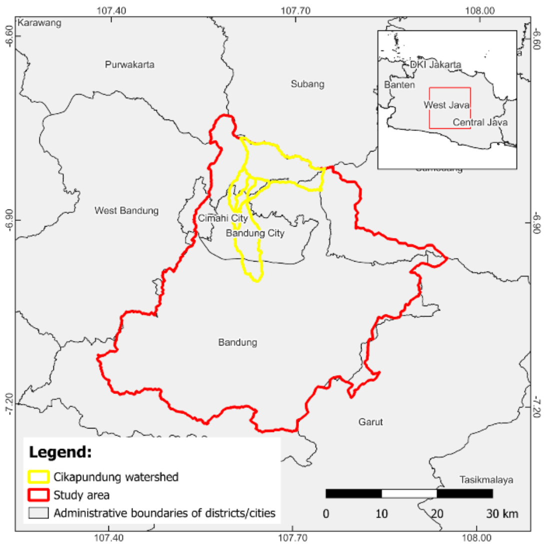

Figure 1. Study area

2.2 Data

Land cover change in the Cikapundung watershed was identified using a combination of Landsat 5 and 8 satellite images to suit the long period of the study. Then, the land cover change factors analyzed by regression were divided into two groups, namely natural and social groups. The natural group includes factors from nature, for which we used temperature and accumulated rainfall data. Meanwhile, the social group includes factors that are heavily influenced by human activities, in this case, population data since 1980 and nighttime light data were used, which also illustrate the development of human life in the area.

Table 1. Data

Dataset	**Format**	**Spatial Resolution**	**Time**	**Source**
Landsat 5	Raster	30 m		
Landsat 8	Raster	30 m	2023	
Population	Raster	100 m	2000 - 2020	World Population
Road	Vector	125 m	2013	Geospatial Information Agency
River	Vector	125 m	2013	Geospatial Information Agency
Nighttime Light	Raster		1992 - 2013	Version 4 DMSP-OLS Nighttime Lights Time Series, NOAA
Temperature and Precipitation	Raster	5 km		Abatzoglou, et. Al., 2018

2.3 Method

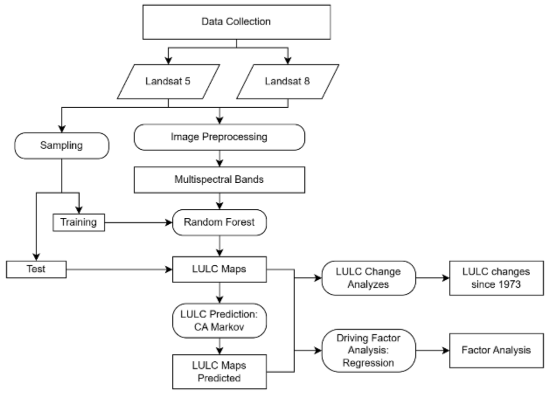

Figure 2. The study general methodology

2.3.1 LULC Classification

In this paper, we perform classification based on Landsat-5 and Landsat-8 images available in the GEE repository. The classification classes in this study are divided into 4 classes, namely open land, built-up area, agricultural vegetation and non-agricultural vegetation, with sample selection through image appearance. Random Forest is the method used in the classification in this study. This method has shown success in the land use and land cover classification process and often achieves the best results compared to other methods (Zafar et al., 2024, Tan et al., 2024, Bwalya Mutale et al., 2024). It is a controlled, nonparametric classification technique that leverages machine learning. The core idea involves generating decision trees, with each tree independently determining the class assignment for each individual pixel. Random Forest refers to a group of ensemble methods that utilize tree-based classifiers {h(x, Θk), k = 1, … } where {Θk} are independent, identically distributed random vectors and x represents the input data (Breiman, 2001). For classification tasks, each tree casts a vote for the most likely class based on the input x, and the final classification is determined by the majority vote of all trees.

2.3.2 LULC Prediction: CA Markov

Markov CA was used to predict the land cover results in 1973. Backward land cover prediction was performed to obtain these results using land cover in 1993 and 1983. CA-Markov is a method that combines the spatial analysis capabilities of Cellular Automata with the statistical strength of Markov chains allowing for the analysis of spatial and temporal dynamics in land cover patterns (Abdelkarim, 2023). Cellular Automata simulates spatial processes by dividing the area into discrete cells, where each cell changes its state based on defined local rules. Markov Chains, on the other hand, offer a probabilistic framework for predicting future states based on current conditions, enabling the modeling of transitions between various land cover types. Together, these methods provide a robust tool for analyzing and forecasting land use and land cover dynamics. In CA-

Markov predictions, the prediction results are based on the range of years between the two land covers used.

2.3.3 Change Detection

Land cover change detection analysis is conducted to monitor and understand the transformations occurring in a specific area over time, with a focus on their attributes, rate, spatial patterns, and driving factors (Tahraoui & Radja Kheddam, 2024; Ruchi Kowarkar et al., 2024). A commonly used technique for this purpose is the post-classification comparison method, which involves examining classified images. To evaluate changes in LULC, transition matrices are generated by comparing thematic maps from various years within the study period.

2.3.4 Factors Affecting LULC Change Analysis

Land cover changes cannot be separated from human and environmental factors. Therefore, a comparison of changes in factors, namely population, nighttime light, temperature, and precipitation accumulation with land cover was carried out using logistic regression using SPSS (Statistical Package for the Social Sciences). This approach used to model the relationship between one or more explanatory variables and a binary or categorical dependent variable. In our context, logistic regression can predict the probability of a specific land cover type occurring based on various explanatory factors. The model estimates coefficients for each explanatory variable, representing the strength and direction of its influence on the probability of a given land cover class. The resulting model follows the equation:

$$ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

where ln is the natural logaritmic, p probability value, α regression constant, and β predictor coefficient.

3. RESULT AND DISCUSSION

3.1 Land Use and Land Cover Classification

LULC classification in the study area has been successfully conducted using the random forest machine learning method for the years 1973, 1999 and 2023. The classified LULC classes are open land, built-up area, agricultural vegetation and non-agricultural vegetation. There are differences in the LULC classification results for each year classified as shown in **Figure 3**.

The LULC classification results show that from 1973 to 2023 the study area was dominated by agricultural vegetation and non-agricultural vegetation classes. Both classes experienced an increase and decrease in area in each classification year. However, both classes have a decreasing trend in area over time as shown in **Table 2** and **Figure 4**.

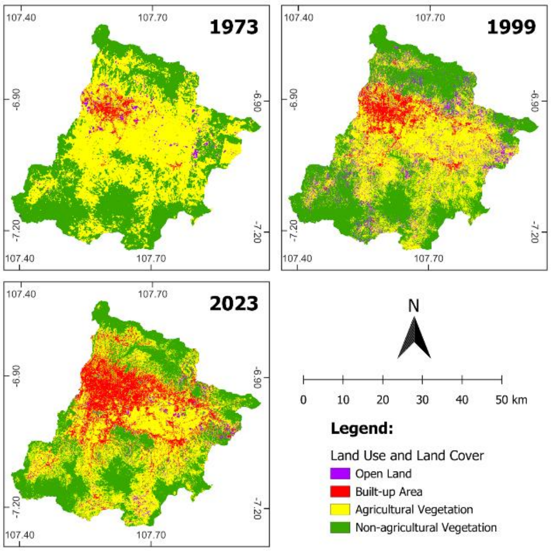

Figure 3. Land use and land cover map of Cikapundung watershed area in 1973, 1999, and 2023

Table 2. Land use and land cover class area

No.	LULC	1973		1999		2023	
		Area (km^2)	%	Area (km^2)	%	Area (km^2)	%
1	Open Land	35.43	1.95	153.93	8.45	52.63	2.89
2	Built-up Area	48.58	2.67	119.50	6.56	309.43	16.99
3	Agricultural Vegetation	938.70	51.55	751.99	41.29	742.11	40.75
4	Non-agricultural Vegetation	798.31	43.84	795.62	43.69	716.85	39.37

Unlike the vegetation class, the built-up area class has a tendency to increase in area over time. In 1973, the built-up land class accounted for 2.67% of the total study area. The peak occurred in 2023, with the built-up area class accounting for 16.99% of the total study area.

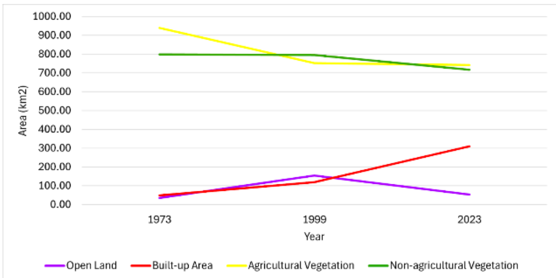

Figure 4. Land use and land cover class area

3.2 Land Use and Land Cover Change

The study also calculated the LULC changes that occurred in the study area. LULC changes were calculated for the years 1973 to 1999 and 1999 to 2023. The LULC changes calculated were LULC changes to the built-up area class and LULC changes to the agricultural vegetation class. This was done to determine the characteristics of LULC change, particularly the built-up area and agricultural vegetation classes, in the study area. The results of the LULC change calculation are shown in **Figure 5**.

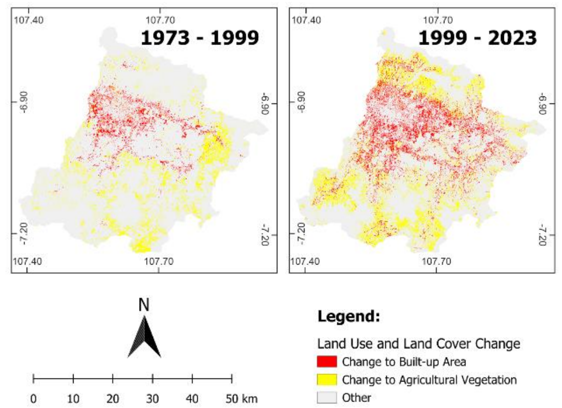

Figure 5. Land use and land cover change in 1973 - 1999 and 1999 - 2023

In general, the change of LULC to built-up area and agricultural vegetation classes has increased over time. The change of LULC to built-up land class is 3.89% and 10.43%, respectively for LULC change from 1973 to 1999 and 1999 to 2023. Meanwhile, the change of LULC to the agricultural vegetation class was 10.07% and 15.81%, respectively for LULC change from 1973 to 1999 and 1999 to 2023. The overall change of LULC to agricultural vegetation class is 12.89%. The results of the calculation of the LULC change are shown in **Table 3**.

Table 3. Land use and land cover change area

LULC Change	1973 - 1999		1999 - 2023	
	Area	%	Area	%
Change to Built-up Area	70.91	3.89	189.94	10.43
Change to Agricultural Area	183.32	10.07	287.96	15.81
Other	1566.80	86.04	1343.14	73.76

3.3 Factors Affecting Land Use and Land Cover Change

This study has assessed the influence of factors causing LULC change (independence variable) using the logistic regression method. The factors analyzed consisted of night time light, population, precipitation, maximum temperature, distance from road, and distance from river. The results of the assessment of factors causing LULC change using the logistic regression method are shown in **Table 4**.

Table 4. Variables in the equation of logistic regression model

Variables or drivers	Change to Built-Up Area			Change to Agriculture Vegetation		
	B	Sig.	Exp(B)	B	Sig.	Exp(B)
Night Time Light (X1)	0.038	0.000	1.039	0.024	0.000	1.025
Population (X2)	0.030	0.000	1.030	-0.024	0.000	0.976
Precipitation (X3)	0.003	0.000	1.003	-0.001	0.000	0.999
Max. Temperature (X4)	0.130	0.000	1.139	-0.305	0.000	0.737
Distance from Road (X5)	0.000	0.000	1.000	0.000	0.000	1.000
Distance from River (X6)	0.000	0.000	1.000	0.000	0.000	1.000
Constant	0.465	0.000	1.591	-2.020	0.000	0.133

The results of the assessment of factors causing LULC changes using logistic regression are in the form of variables in the logistic regression model equation. The resulting variables are the coefficient (B), significance (Sig.), and odds ratio (Exp(B)) as shown in **Table 4**. The coefficient (B) value indicates the strength and direction of the relationship between the independent variable and the dependent variable. The significance (Sig.) value indicates the significance of the effect of the independent variable in influencing the dependent variable. Meanwhile, the odds ratio (Exp(B)) shows the relative change in the odds of the dependent variable due to a one-unit change in the dependent variable. The equation of the logistic regression model of the independent variable and the dependent variable for the change of LULC to built-up land class and agricultural vegetation are shown in the following equations, respectively.

$$\ln\left(\frac{p}{1-p}\right) = 0.465 + 0.038\mathrm{X}_1 + 0.030\mathrm{X}_2 + 0.003\mathrm{X}_3 + 0.130\mathrm{X}_4 + 0.000\mathrm{X}_4 + 0.000\mathrm{X}_6$$

$$\ln\left(\frac{p}{1-p}\right) = -2.020 + 0.024\mathrm{X}_1 - 0.024\mathrm{X}_2 - 0.001\mathrm{X}_3 - 0.305\mathrm{X}_4 + 0.000\mathrm{X}_5 + 0.000\mathrm{X}_6$$

Based on the results of the logistic regression assessment, each independent variable has a diverse influence on the dependent variable. The logistic regression analysis shows that factors such as nighttime light, population, precipitation, and maximum temperature significantly drive these changes, both of LULC change to built-up area and LULC change to agricultural vegetation. However, distance from roads and rivers showed no significant influence on LULC change.

4. CONCLUSION

The study area has been classified into four LULC classes: open land, built-up area, agricultural vegetation, and non-agricultural vegetation. The LULC classes in the study area are dominated by the agricultural and non-agricultural vegetation classes which have a trend to decrease in area over time. Meanwhile, the built-up area class has an increasing trend. From 1973 to 2023, there has been a 14.32% change in LULC to built-up area and a 12.89% change to agricultural vegetation. Logistic regression has been used to identify the influence of factors that may cause LULC change. Factors such as night time light, population, precipitation, and max temperature have a significant influence on LULC change. Whereas, factors such as distance from roads and distance from rivers do not influence significantly on LULC change.

5. ACKNOWLEDGEMENTS

This research was funded by LPPM-ITB under the Program Riset ITB 2024; 1583/IT1.C01/SK-TA.00/2024 grant scheme attributed to the Geospatial Information Science and Technology Research Group (KK-STIG).

REFERENCES

Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, 2018, Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data 5:170191, doi:10.1038/sdata.2017.191

Abdelkarim, A. (2023). Monitoring and forecasting of land use/land cover (LULC) in Al-Hassa Oasis, Saudi Arabia based on the integration of the Cellular Automata (CA) and the Cellular Automata-Markov Model (CA-Markov). Geology, Ecology, and Landscapes, 1–32. https://doi.org/10.1080/24749508.2022.2163741

Agaton, M., Setiawan, Y., & Effendi, H. (2016). Land Use/Land Cover Change Detection in an Urban Watershed: A Case Study of Upper Citarum Watershed, West Java Province, Indonesia. Procedia Environmental Sciences, 33, 654–660. https://doi.org/10.1016/j.proenv.2016.03.120

Arno Adi Kuntoro, Anton Winarto Putro, Muhammad, & Suardi Natasaputra. (2017). The effect of land use change to maximum and minimum discharge in Cikapundung River Basin. Nucleation and Atmospheric Aerosols. https://doi.org/10.1063/1.5011621

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324

Bwalya Mutale, Neel Chaminda Withanage, Mishra, P. K., Shen, J., Abdelrahman, K., & Fnais, M. S. (2024). A performance evaluation of random forest, artificial neural network, and support vector machine learning algorithms to predict spatio-temporal land use-land cover dynamics: a case from lusaka and colombo. Frontiers in Environmental Science, 12. https://doi.org/10.3389/fenvs.2024.1431645

Fardani (2020). Landuse change prediction model based on Cellular Automata (CA) method in Bandung City. Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1469/1/012030

Liu L, Yu S, Zhang H, Wang Y, Liang C. Analysis of Land Use Change Drivers and Simulation of Different Future Scenarios: Taking Shanxi Province of China as an Example. Int J Environ Res Public Health. 2023 Jan 16;20(2):1626. https://doi.org/10.3390/ijerph20021626.

Ruchi Kowarkar, Gupta, D., Jatin Karkera, & Kulkarni, S. (2024). Exploring Change Detection in Classified Land Cover and Land Use Imagery: A Comprehensive Analysis. 2024 Second International Conference on Data Science and Information System (ICDSIS). https://doi.org/10.1109/icdsis61070.2024.10594277

Shang S., Cui T., Wang Y., Gao Q., Liu Y. (2024). Dynamic variation and driving mechanisms of land use change from 1980 to 2020 in the lower reaches of the Yangtze River, China. Frontiers Environmental Science, Volume 11. https://doi.org/10.3389/fenvs.2023.1335624

Sunandar, Y. P. (2024, February 19). Mengenal DAS Citarum Hulu. Walungan. https://walungan.org/2024/02/19/mengenal-das-citarum-hulu/

Tahraoui, A., & Radja Kheddam. (2024). LULC Change Detection Using Combined Machine and Deep Learning Classifiers. 2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP), 403–408. https://doi.org/10.1109/atsip62566.2024.10638952

Tan, Y.-C., Duarte, L., & Teodoro, A. C. (2024). Comparative Study of Random Forest and Support Vector Machine for Land Cover Classification and Post-Wildfire Change Detection. Land, 13(11), 1878–1878. https://doi.org/10.3390/land13111878

Yonik Meilawati Yustiani, & Leony Lidya. (2016). Towards an Information System of Modeling and Monitoring of Cikapundung River, Bandung, Indonesia. Procedia Engineering, 154. https://doi.org/10.1016/j.proeng.2016.07.490

Zafar, Z., Zubair, M., Zha, Y., Fahd, S., & Ahmad Nadeem, A. (2024). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. The Egyptian Journal of Remote Sensing and Space Sciences, 27(2), 216–226. https://doi.org/10.1016/j.ejrs.2024.03.003