The Diminishing Green Land in the Buffer City Area of Jakarta

Venezia Valen Susilo* and Dyah Erny Herwindiati

Faculty of Information Technology, Tarumanagara University, Jakarta, Indonesia
Email: valenvenezia@gmail.com (V.V.S.); dyahh@fti.untar.ac.id (D.E.H.)
*Corresponding author

***Abstract*—Jakarta got its water from Mt. Salak in Bogor, land it is not directly adjacent. From Mt. Salak to Jakarta, the water has to pass a few other regencies like Bogor, Depok, and South Tangerang. That means whatever changes happen in those areas will affect how the water reaches Jakarta. Therefore, a system is needed to track how the land changes in those areas, and one of them is through satellite images, Landsat 8. The images taken will give a rough visual of the land, but as data, it has to be in numbers. A method called Gradient Boosting Regression helped with that. This method can build a model that classifies every pixel of the images and, in this research, green, partial green, and impervious. This model has an accuracy of 99.3349% for the training data and 99.1658% for the validation data and took 13.91376 seconds to complete. From this model, there is a trend where the green area keeps decreasing through the years, and the district with the highest area of land-use change from green to impervious is Sukamakmur in Bogor.**

***Keywords*—Jakarta, green area, Landsat, gradient boosting regression, image classification**

I. Introduction

The need for land in the JABODETABEK (Jakarta, Bogor, Depok, Tangerang, Bekasi) continues to grow with the population and economic growth. That leads Jakarta, the capital city of Indonesia and the center of the country's economy, to have less and less land remaining. On top of that, developing house clusters and high-rise buildings also have become a trend in recent years. The need for land forces the said developers to change the use of those remaining lands from green lands to impervious land [1]. Land-use change itself is a change of function of either partially or the whole area from its original function to another that can affect the environment negatively [2].

According to Law Number 26/2007 about Spatial Planning, a minimum of 30% of the total area must be designated as green open space [3]. However, according to statistics in 2020, Jakarta only has 9.4% of green area, which is far from the supposed number [4].

The land use change for the sake of development cannot continue for a long time as there is not much space left and the side effect of the said process. Green space is essential in a city for its role as the city's lung and water absorption. The diminishing of the green space area causes the decrease of oxygen levels in the air and water absorption area, which can cause flooding [5].

Diminishing the water absorption area also means fewer water resources in Jakarta. The water must first flow through several cities like Bogor City, Depok, and South Tangerang through the Ciliwung River to reach Jakarta from Mt. Salak in Bogor. Therefore, whatever changes happen in those areas will also cause a change in how the water reaches Jakarta. If some problems arise in the areas in question, Jakarta, as the destination point, will be the one who gets the most significant impact.

Because of the reasons above, there is a need for a way to monitor the land use changes from green areas to impervious land or vice versa. That way, the researcher can see if a development project is good to be done or if it will bring other impacts on the environment and the water flow.

This paper will discuss a system that calculates and maps the Depok, Bogor, and South Tangerang districts from 2014 to 2020. The land will be classified into three classes: green land, partial green land, and impervious land. The classification process will use the Gradient Boosting Regression method.

This system aims to give information about the land use changes that happened in Bogor, Depok, and South Tangerang. This information is desired to help the parties related to development in those areas to mitigate the impacts of the land use change from the development. From the mapping that the system will provide, it is hoped that related parties can monitor which area is still green and which is not.

Previously, there was research on this topic by Devin Budi in 2021 with "Land Use Change Using Least Absolute Shrinkage and Selection Operator Regression in Jakarta's Buffer Cities" [6]. The research achieved a LASSO LARS model with an accuracy of 81.14% for Landsat 7 and 77.52% for Landsat 8 images.

Another research is from Juni Handoko in "Gradient Boosting Tree for Land Use Change Detection Using Landsat 7 and 8 Imageries: A Case Study of Bogor Area as Water Buffer Zone of Jakarta" [7]. This research achieved an F1 Score of 69.04% with 90 trees.

A. Collecting Data

The help of remote sensing is needed to keep track of land use change. It is necessary to see how the land is, but

Manuscript received February 5, 2023; revised April 7, 2023; accepted June 28, 2023; published January 23, 2024.

more than a simple camera is needed. Remote sensing is a technology to get information about an object, area, or phenomenon through an analysis of data that was taken with a tool that does not correlate with the object of interest. The goal of remote sensing is to capture the data and information from the imageries of the earth that were taken with an artificial censoring tool [8].

The data used in this research is Landsat 8 imageries pulled from the site of the United States Geological Survey (USGS, https://earthexplorer.usgs.gov/ that were taken from 2014 to 2020 for every 16 days around the area of Bogor, Depok, and South Tangerang, which is on path 122 and row 64–65. From the 11 bands that Landsat 8 takes, the ones that will be used in this research are from band 2 – band 7. Band 2–4 are chosen to create the image as it was by combining band 4(red), band 3(green), and band 2(blue). Band 5 is chosen because it is the band that captures the infrared light that is used to check the greeneries and plants. Bands 6 and 7 are chosen as these are the bands that identify the wetness of a land. Each pixel that Landsat 8 takes represents a 30 m × 30 m area in real life. Therefore, a pixel represents 900m^2 [9].

The scope of the area consists of Depok with 200.3 km^2, Bogor Regency with 2,986 km^2, Bogor with 118.5 km2, and South Tangerang with 147.2 km^2. The total area of the regencies is 3,452 km^2. An example of the imagery can be seen in Fig. 1.

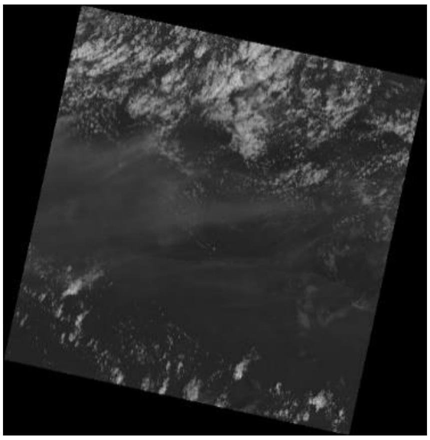


Fig. 1. February 1, 2014 Landsat 8 band 2 image.

Landsat 8 takes the image with a size of 16 bits, which is why the images need to be pre-processed to transform them into 8 bit images so they will have the data range of 0–256. Images that have been transformed will be cropped to fit the shape of the districts' borders so that every district will consist of 6 images from band 2 to band 7. There will be a total of 59 sets of images, which are 41 districts in Bogor, 11 districts in Depok, and seven in South Tangerang. Those images will be cropped with a software called ENVI, and the border shapefiles were provided by Dr. Sulaiman from BPPT (Badan Pengkajian dan Penerapan Teknologi). The images cut according to the district border shapefile will be used as the test dataset. An example of the pre-processing result can be seen in Fig. 2.

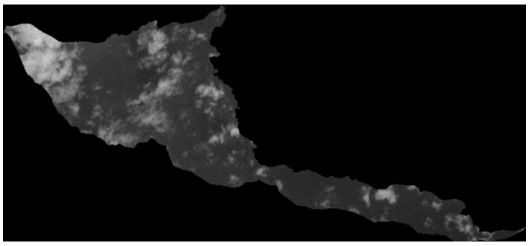


Fig. 2. Pre-processing result of District Cigombong in Bogor.

There is a difference between the cropping method for the training dataset and the test dataset. Before cropping, the classified train data must be determined by choosing the areas for each class. Whether the area is green, partially green, or impervious can be determined by looking at the real satellite imageries through the Google Earth Pro software and turning it into polygon files. Use ArcMap software to convert the polygon files to shapefiles. The transformed 8 bits of Landsat 8 imageries will be cropped by using those shapefiles to make the training dataset. The result of the pre-processed train data can be seen in Fig. 3.

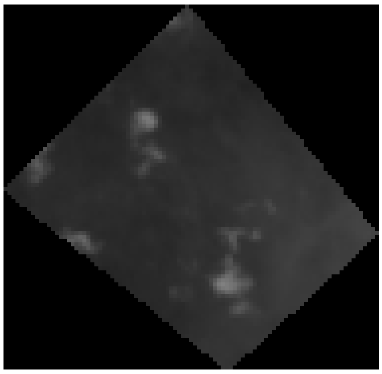


Fig. 3. Example of train green area image.

B. Algorithm

The method used in this research is a classification method with Gradient Boosting Regression. Gradient Boosting Regression is one of the weak-learner machine learning that is used to build a prediction or regression model.

The built model will be used to classify the areas of Bogor, Depok, and South Tangerang districts into either green, partial green, or impervious. This method is used for several reasons, like its flexibility with many loss functions and the main reason being the iteration process. The method will enhance the model in every iteration based on the residual rate from the previous iteration. This way, the residual rate will decrease along with the iteration.

The loss function used in this paper is Logistic Function; some call it 'deviance loss' [10]. The logistic function is obtained by looking for a log value from the probability of a datum classified as a specific class. The probability equation is:

$$p_k(x) = \frac{e^{f_k(x)}}{\sum_{l=1}^{K} e^{f_l(x)}} \quad (1)$$

where:
k = class k
K = number of classes
$f(x)$ = prediction tree value

Every iteration will reduce the residual rate by making a regression tree that will depend on the residual rate from the previous iteration. Below is the equation to determine the leaves value [11].

$$\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} r_{ikm}}{\sum_{x_i \in R_{jkm}} |r_{ikm}|(1-|r_{ikm}|)} \quad (2)$$

with:
γ_{jkm} = value of leaf j of class k in iteration m
r_{ikm} = Residual of data I of class K in iteration m

The algorithm of gradient boosting regression is:

1. Initialize $f_{k0}(x) = 0, k = 1, 2, \dots, K$.
2. For $m = 1$ to M:
   a. Count the probability value with equation (1).
   b. For $k = 1$ to K:
      i. Calculate $r_{ikm} = y_{ik} - p_k(x_i), i = 1, 2, \dots, N$.
      ii. Build regression tree according to the target $r_{ikm}, i = 1, 2, \dots, N$. Make leaves according to terminal $R_{jkm}, j = 1, 2, \dots, J_m$.
      iii. Count every leaf value with equation (2).
      iv. Update $f_{km}(x) = f_{k,m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jkm} I(x \in R_{jkm})$.
3. Output $\hat{f}_k(x) = f_{kM}(x), k = 1, 2, \dots, K$.

C. Design

The flowchart of the system can be seen in Fig. 4.

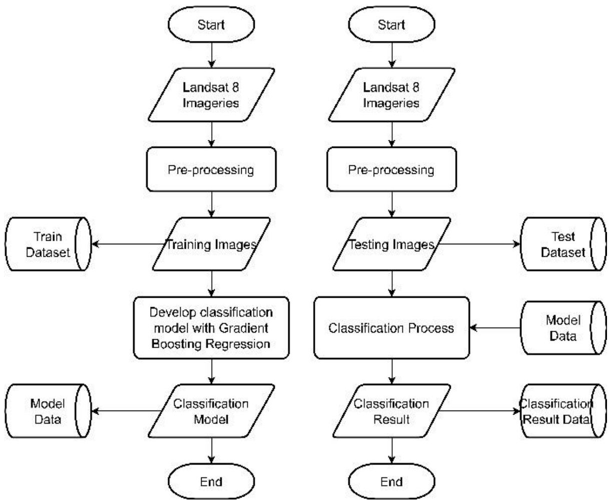


Fig. 4. Process flowchart.

The Landsat 8 images of the area around Bogor, Depok, and South Tangerang from 2014 to 2020 are downloaded as the raw data for training and test datasets. With the Google Earth Pro software, search the areas for each class, green, partially green, and impervious, as the training dataset. Once done, do the pre-process as what is explained in II.A. Collecting Data. The images will be converted to 8 bits images and cropped according to the shapefile of the train data from Google Earth Pro.

The training dataset consists of 5 sets of green area images, five partially green area images, and five impervious land images. The training dataset is divided into 2, the training dataset and the validation dataset, with a ratio of 35% validation and 65% training. The training dataset is used to build the Gradient Boosting Regression classification model, and the validation dataset is used to test the model's accuracy. The train data have 15,589 pixels of green area, 29,374 pixels of partially green areas, and 18,484 pixels of impervious land. In comparison, the validation data have 8,417 pixels of green areas, 15,707 pixels of partially green areas, and 10.041 pixels of impervious land. The example of the train data can be seen in Table I.

TABLE I. EXAMPLE OF TRAIN DATA

B 2	B 3	B 4	B 5	B 6	B 7	Class
49	50	45	128	113	88	Green
51	54	48	135	117	92	Green
59	60	56	132	118	93	Green
64	64	75	102	134	147	Partial Green
64	63	73	106	136	144	Partial Green
61	59	70	101	129	138	Partial Green
56	70	86	129	137	142	Impervious
41	57	65	118	107	98	Impervious
44	59	64	78	72	72	Impervious

The training process will produce a model that will be used to classify the districts' images. Once a model is built, it will be tested for accuracy with the confusion matrix to calculate the F1 Score. Several models will be made to find the best one, either in accuracy or efficiency.

Each pixel of the testing image will be classified with the chosen model and mapped according to its predicted class. The mapped image will be compared to another mapped image of the same district taken at a different time to calculate the land-use change.

D. Evaluation Method

Evaluation method that is used in this research is the F1% value. The F1% value is achieved by determining the accuracy percentage from a confusion matrix. The confusion matrix is one of the many evaluation methods for classification models by comparing the actual value with the prediction value.

The confusion matrix is divided into four areas which are True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN). True positives represent the amount or percentage of 'yes' data predicted as 'yes'. True negatives represent the amount or percentages of 'no' data that is correctly predicted as 'no'. False positives represent the amount or percentages of 'no' data that is falsely predicted as 'yes'. False negatives represent the amount or percentages of 'yes' data that is predicted as 'no'. The model's accuracy can be measured by calculating the F1 Score. However, to be able to do that, the value of precision and recall have to be determined first. Precision is a value of how many correctly 'yes' data are predicted out of all predicted 'yes' data, while recall is a value of how many correctly 'yes' data out of all real 'yes' data. The equation can be seen in equation (3) to equation (5). An example of a confusion matrix can be seen in Fig. 5 [12]:

$$precision = \frac{TP}{TP+FP} \tag{3}$$

$$recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} \tag{5}$$

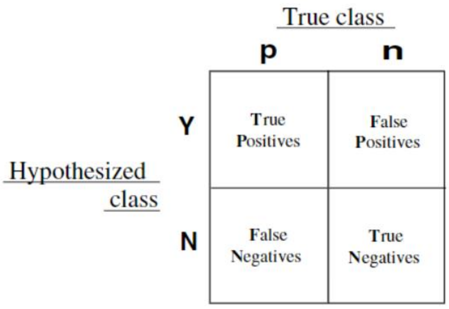


Fig. 5. Confusion matrix.

II. Results and Discussion

In the training process, it is impossible to get the best model and result in one try, hence the experiments. The training process is done many times to find the best model that can be used in the testing process. The trials are done by trying several times by changing the parameters like the amount of tree. All models are tested for their accuracy and the time it takes to complete building. The chosen model has to be the one with the highest accuracy but the fastest time. The result of the testing can be seen in Table II.

TABLE II. Training Models Result

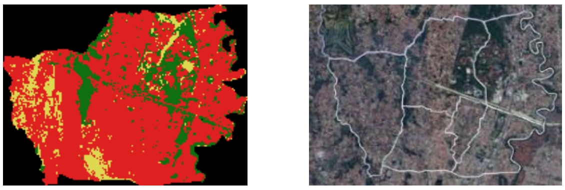
Number of trees	Accuracy (Training data)	Accuracy (Validation data)	Time (second)
10	0.978486	0.977872	3.08069
20	0.985484	0.984721	5.755840
30	0.989456	0.988204	8.3177196
50	0.993349	0.991658	13.91376
60	0.994641	0.992683	16.19372
85	0.996611	0.993590	23.31680
100	0.997478	0.994146	25.9046752

From Table II, it can be seen that the more trees used in a model, the higher the accuracy, but the time it took to build the model is also longer. Considering the accuracy rate and the building time, the model with 50 trees is the best out of the others. The 50-tree model can reach an accuracy higher than 99%, and to be exact, 99.3349% for train and 99.1658% for validation within a bit less than 14 seconds. The value of precision, recall, and F1 Score can be seen in Table III.

TABLE III. Model Accuracy Table

Class	Precision	Recall	F1 Score
Green	0.9929	0.9907	99.18 %
Partial Green	0.9916	0.9947	99.31 %
Impervious	0.9907	0.9878	98.92 %

After the model building is done, the model is used to run on the test dataset. Landsat 8 images that is done with the whole pre-processing steps are mapped with the said model. Pixels that are classified as green area will be colored green. Pixels that are classified as partially green will be colored light green. Lastly, the impervious pixels are colored in red. The comparison of the mapped image with the ground truth can be seen in Fig. 6.

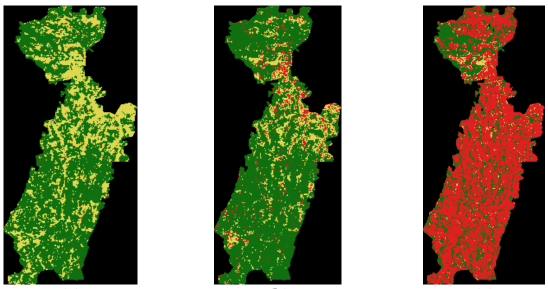


(a) (b) (c)

Fig. 6. Comparison of the mapped Beji (2020) to ground truth image from Google Earth: (a) mapped image, (b) Google Earth image, (c) overlaid image.

From Fig. 6, it can be seen that the result of the mapping can represent the ground truth type well. The green land from the Google Earth image is also classified as green, and land with buildings is classified as impervious in the mapped result.

To map Beji, which has a size of 175×130 pixels with the model, it took 1.8 seconds. The time it took to map is different for each district, depending on the image size. The longest district is Rumpin in Bogor regency, with 377×718 pixels. The model took 3.8 seconds to map Rumpin. Some mapping results can be seen in Fig. 7 to Fig. 9.



(a) (b) (c)

Fig. 7. Classification result of Dramaga, Bogor: (a) 2015, (b) 2018, (c) 2020.

(a) (b) (c)

Fig. 8. Classification result of Beji, Depok: (a) 2015, (b) 2018, (c) 2020.


(a) (b) (c)

Fig. 9. Classification result of North Serpong, South Tangerang: (a) 2015, (b) 2018, (c) 2020.

TABLE IV. CLASSES' AREAS OF DRAMAGA, BEJI, AND SOUTH TANGERANG IN 2015, 2018, AND 2020

Year	District	Green Area (m^2)	Partial Green Area (m^2)	Impervious Area (m^2)
2015	Dramaga	17,334,900	8,096,400	192,600
	Beji	2,799,000	1,318,500	10,466,100
	North Serpong	5,182,200	8,366,400	8,573,400
2018	Dramaga	19,677,600	4,147,200	1,799,100
	Beji	896,400	34,200	13,653,000
	North Serpong	1,336,500	17,100	20,768,400
2020	Dramaga	6,546,600	765,900	18,297,900
	Beji	2,290,500	1,330,200	10,962,900
	North Serpong	857,700	663,300	15,215,400

It can be seen from Table IV that the green area from 2020 is much less than what it was in 2015 for all three districts. This data showed the diminishing green land in these regencies over the years.

The mapped image will be compared to the other mapped image of the same district but taken at a different time to see the land-use change. For example, the mapped image of the 2015 Beji will be compared to the mapped 2018 Beji image. The method is by creating a new image by comparing each pixel to see if there is a difference between the two images.

If there is no difference, the pixel of the new image will be colored in white. The change from the green pixel to the partial green pixel will be colored beige. The change from the green pixel to the impervious pixel will be colored in red; from the partial green pixel to the green pixel will be colored in light green; from partial green to impervious in pink, impervious to green in dark green, and impervious to partial green in orange. An example of land-use change mapping of Beji can be seen in Fig. 10, with more detail in Table V.

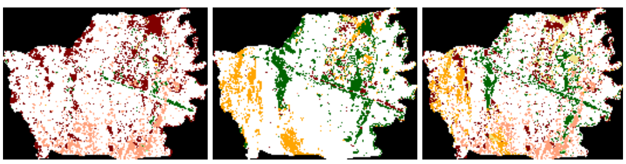

(a) (b) (c)

Fig. 10. Beji land-use change map: (a) 2015–2018, (b) 2018–2020, (c) 2015–2020.

TABLE V. AREA OF BEJI'S LAND-USE CHANGE

Year 1	Year 2	Original Land Type	Changed Land Type	Area (m^2)
2015	2018	Green	Partial Green	27,900.00
			Impervious	2,160,900.00
		Partial Green	Green	17,100.00
			Impervious	1,301,400.00
		Impervious	Green	269,100.00
			Partial Green	6,300.00
		No change		10,800,900.00
2018	2020	Green	Partial Green	90,000.00
			Impervious	383,400.00
		Partial Green	Green	0
			Impervious	6,300.00
		Impervious	Green	1,867,500.00
			Partial Green	1,212,300.00
		No change		11,024,100.00
2015	2020	Green	Partial Green	456,300.00
			Impervious	1,232,100.00
		Partial Green	Green	28,800.00
			Impervious	1,187,100.00
		Impervious	Green	1,151,100.00
			Partial Green	771,300.00
		No change		9,756,900.00

As seen from Table V, the most significant change in Beji was from green to impervious from 2015 to 2018, with 2,160,900.00 m^2. Beji has a total area of 14,560,000.00 m^2; this means 15% of Beji changed from green areas to impervious land. The top 10 area where the land use change from green to impervious out of all 59 districts from 2015 to 2020 can be seen in Fig. 11.

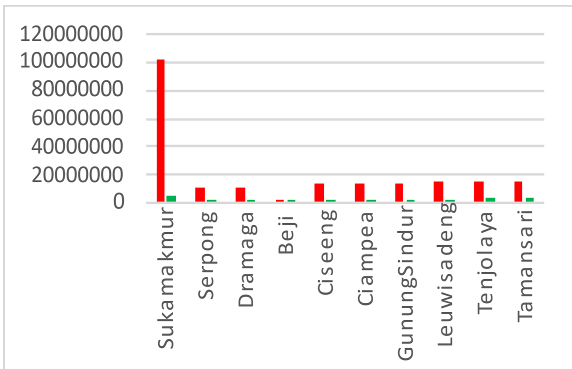


Fig. 11. Chart of land-use change from green to impervious.

From Fig. 11, the district with the most significant land-use change is Sukamakmur District in Bogor Region, with 101,761,200.00 m^2. With red bar representing the land-use change from green to impervious and green bar for impervious to green, it was apparent that there is more red than green. From this chart, it can be inferred that there is far more land-use change to impervious land. The comparison between the green area in 2015 and 2020 can be seen in Fig. 12.

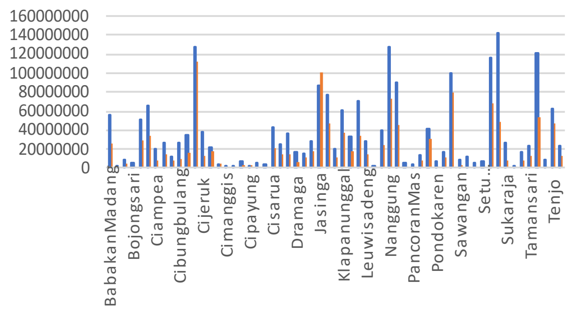


Fig. 12. Green area in 2015 vs 2020.

From Fig. 12, with the blue bar representing the green area in 2015 and the orange bar represents the green area in 2020, it was crystal clear that most districts have a higher blue bar than their corresponding orange bar. It meant that the green area of those districts was much higher in 2015 than in 2020. From that data, again, Sukamakmur in Bogor is the district that lost the green area the most, with a total of 92,642,400.00 m^2.

With this piece of information, land observers or local government can track down why this considerable change happen and can also analyze the effect of this change. It is hoped that this data can push for a better environment policy.

III. Conclusion

From the training and testing result, it can be inferred that Gradient Boosting Regression is an excellent method for building a model for classifying Landsat 8 imageries. The best model is achieved by building it with 50 trees as the model can successfully classify 99.18% of green area, 99.31% of partial green area, and 98.92% of impervious land and only needs 13.9 seconds to be built. Comparing this method to others method like LASSO LARS model with accuracy rate of 77.52% for Landsat 8 images, these numbers are much higher.

This system is hoped so it can help parties that are involved in the regencies' development can track the land-use change to maintain the area better either for the citizens or the environment by keeping an eye on how much green area is left.

Like what happened to Beji from 2015 to 2018, where 15% of the whole area turned from green to impervious, it shouldn't be a trend as there is not much green area left, as can be seen in Fig. 8 as of 2020.

Fig. 11 and Fig. 12, the charts showed how the green area of most Bogor, Depok, and South Tangerang districts kept diminishing from what it was in 2015. Sukamakmur is the district with the highest land-use change rate, with 101,761,200.00 m^2 of its green space turned into impervious land, and the one that lost its green area the most out of the three regencies from 2015 to 2020, with the total loss reached 92,642,400.00 m^2.

The pre-process just converts the 16-bit images to 8-bit and crops them. However, there is another challenge in collecting the images, which is the cloud. The satellite takes the imageries at regular intervals, so it does not discern when is the right time to capture the image and when it is not. That is why the images will be covered in the cloud, especially in rainy seasons. The areas that are covered with cloud won't be able to be correctly classified with the model. The research would produce a better result if the images were through a cloud-removal process in the pre-processing step.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Venezia Valen Susilo did the experiment, computation and analyzing the result; Dyah Erny Herwindiati conducted the research and topic of the paper; all authors worked on the paper and had approved the final version.

References

[1] D. E. Herwindiati, J. Hendryli, and L. Hiryanto, "Impervious surface mapping using robust depth minimum vector variance regression," *European Journal of Sustainable Development*, 2017.

[2] A. Nurpita, L. Wihastuti, and Y. Andjani, "The impact of land conversion on the food security of farming households in Temon sub-district, Kulon Progo district," *Gama Society Journal*, p. 105, January 2018.

[3] W. Brontowiyono, R. Lupiyanto, D. Wijaya, and J. Hamidin, "Urban heat islands mitigation by Green Open Space (GOS) canopy improvement: A case of Yogyakarta Urban Area (YUA), Indonesia," *International Journal of Technology*, pp. 207–214, 2011.

[4] I. Muhtarom. (4 August 2021). The law requires green open space to be 30 percent of the city area, has it been fulfilled? [Online]. Available: https://tekno.tempo.co/read/1437953/uu-wajibkan-ruang-terbuka-hijau-30-persen-dari-luas-kota-sudahkah-terpenuhi

[5] CiptaKarya. (15 August 2021). 6 benefits of green open space. CiptaKarya. [Online]. Available: http://sim.ciptakarya.pu.go.id/p2kh/knowledge/detail/6-manfaat-ruang-hijau-terbuka

[6] D. Budi, D. E. Herwindiati, and J. Hendryli, "Land use change using least absolute shrinkage and selection operator regression in Jakarta's buffer cities," in *Proc. 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2021.

[7] J. Handoko, D. E. Herwindiati, and J. Hendryli, "Gradient boosting tree for land use change detection using Landsat 7 and 8 imageries: A case study of Bogor area as water buffer zone of Jakarta," *IOP Conference Series: Earth and Environmental Science*, no. 581, 2020.

[8] D. Yusuf and A. S. Rijal, *Remote Sensing*, Gorontalo: UNG Press, 2017.

[9] C. Loyd. (8 January 2022). Landsat 8 bands. [Online]. Available: https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-bands/

[10] J. Brownlee. (30 August 2021). Loss and loss functions for training deep learning neural networks. [Online]. Available: https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/

[11] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 2nd ed., New York: Springer, 2009.

[12] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, no. 27, pp. 861–874, 2006.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.