International Journal of Geospatial and Environment
DOI: 10.31315/ijgev.0203

Stochastic Gradient Boosting Algorithm For Land Use Change Detection Using Multi-temporal Landsat 5 TM In Yogyakarta City, Indonesia

Sintha Prima Widowati Gunawan[1,2,*], Takanori Matsui[1], Takashi Machimura[1]

[1] *Department of Sustainable Energy and Environmental Engineering, Osaka University, 2-1Yamadaoka, Suita, Osaka 565-0871, Japan.*
[2] *Department of Environmental Engineering, University of Pembangunan Nasional "Veteran" Yogyakarta, Jl. Ring Road Utara No.104, Sleman, Yogyakarta 55283, Indonesia*
* *Corresponding author* : sinthawidowati.upnyk@gmail.com

Abstract

The advanced technology in remote sensing and geographic information system has facilitated a great deal to land management through land use/land cover (LULC) change detection using spatial and temporal data to analyze the dynamic conversions on the land surface in two or more periods of time. Classification is the main exercise to present this information to the decision makers. A hybrid bagging-boosting machine learning algorithm was used to generate a binary classifier for separating Urban and Non-Urban classes. A high-resolution image would be such a promising opportunity to achieve high classification accuracy, but issues of availability and feasibility have forced the analyst to settle for medium-resolution spatial data with an imbalanced dataset to detect LULC changes of urban agglomeration area in Yogyakarta city, Indonesia, in the years 1999, 2005, and 2011. The result showed that stochastic gradient boosting algorithm was succeeded in building one robust classifier model using LANDSAT-5TM 2005 with an overall accuracy of 0.76 and ROC-AUC value of 0.83. The replicability of classifier was confirmed by an agreement between the predicted class and the reference data from Statistics Indonesia, which showed root mean square error (RMSE) of 9.9% and R2 of 0.91, indicating sufficiently good accuracy for areal integrated multi-temporal urbanization monitoring.

Keywords: boosting algorithm, land use, change detection, landsat, yogyakarta

1. Introduction

Globally, the urban population in developed and industrializing countries is rapidly expanding to the periphery, shifting the development pressure to local resources of peri-urban areas, and there are still gaps in the integrated planning approaches to address it (Cohen, 2004; Geneletti et al., 2017). Goal 11 of the United Nations Sustainable Development Goals (SDGs) has called all countries to "make cities and human settlements inclusive, safe, resilient and sustainable" targeted to "support positive economic, social, and environmental links between urban, peri-urban and rural areas by strengthening national and regional development planning" (UN, 2020). Land Use/Land Cover (LULC) change detection can provide familiarity to urban and peri-urban trade-offs, assisting local government and decision makers in implementing sound-sustainable planning and management of its land and natural resources (Geneletti et al., 2017; Tewabe and Fentahun, 2020).

The advanced technology in remote sensing and geographic information systems has facilitated the LULC change detection with satellite images to analyze dynamic conversions on the land surface in two or more periods of time (Richards, 2013). One of most popular change detection analysis techniques is post comparison that compares satellite images in two or more periods of time after they have been individually classified; thus, image classification is critical in LULC change detection analysis (Heydari and Mountrakis, 2018; Mishra et al., 2017). Medium spatial resolution (25–150 m) imagery data such as LANDSAT products [USGS], which can capture the land cover in relatively large areas from 1972 to the present, has emerged as the most favorable imagery data for pixel-level sequential analysis in describing trends and phenomenon on the land surface (Radočaj et al., 2020; Rocchini et al., 2017). Per-pixel classification depends on spectral information featured in the image's pixel, consisting of several channels/bands (visual, infra-red, thermal to panchromatic wavelength) captured from an electromagnetic signal of the Earth's surface (Lillesand et al., 2015; Lu and Weng, 2007). Each value is then converted to a positive numeric digit using analog to digital signal conversion, namely digital number (DN) of 8 or more binary digits that becomes variables assigned to a pixel vector (Lillesand et al., 2015). DN value is most likely altered from one image to another due to changes on the land surface, differences in radiation, atmospheric conditions, and sensor calibration. Thus it is statistically acceptable as an indicator of change in certain land surface area (Mishra et al., 2017).

In order to reduce the complication of interpreting results, the supervised classification (SC) technique which works with probability density function, is a better option because it incorporates sufficient ground truth information before classification and can quantify the accuracy promptly using a statistical by-product, namely confusion matrix to assess the characteristics of case allocations from the classification, such as overall accuracy (OA), precision and recall, F-measurement, and kappa coefficient of agreement (Foody, 2002; Kumar and Sahoo, 2012; Richards, 2013).

The automated decision-making technology in data science statistics called machine learning has facilitated SC effectively, especially when involving bigger data. It employs a set of statistical tools and methods to allow a computer program to learn from some observations/cases to identify patterns and perception, involving more observations to boost the decision-making power of the machine learning algorithm (Gutierrez, 2015). There are three major steps in SC with machine learning methods: preparing the training dataset, selecting the classification algorithm, and finally delivering the result and classification accuracy (Afaq and Manocha, 2021; UKEssays, 2018). In LULC classification with machine learning, a training dataset (e.g., a set of satellite data paired with observed LULC type in a particular area of interest) will be used to build a classifier model so that the classifier performs well on the training dataset as well as on the test datasets (e.g., datasets from other satellite images that are not used to train the classifier) (Gutierrez, 2015). Selecting an algorithm for LULC classification using a medium spatial-resolution image is challenging because it must encompass the available training pixels in order to calculate the optimum hyperparameters to perform the best separation of the spectral values and ensure that the algorithm is reproducible for time-series land cover change analysis to support sound-sustainable land use planning and management (DeFries and Chan, 2000; Kumar and Sahoo, 2012; Lu and Weng, 2007; Richards, 2013).

A variety of machine learning techniques and their characteristics for LULC classification from time-series imagery data are discussed in (Gómez et al., 2016; Lu and Weng, 2007). Parametric algorithms, such as maximum likelihood estimation (MLE), can produce high accuracy classification result in less time-consuming and hassle-free procedures as reported by Kaliraj (Kaliraj et al., 2017), showing that high OA using ETM+ data (81.16%) and TM data (77.52%) facilitated with ERDAS Imagine (HEXAGON™). Mohajane et al. (Mohajane et al., 2018) showed that high OA using TM data (99.9%) and ETM+ data (99.8%) with an open source software, Quantum GIS ver.3.0.0. However, MLE is a fixed linear function, and it is prone to produce noisy results when the training field has complex land cover types; thus, it cannot be reliable to handle bigger size data (Lu and Weng, 2007; Richards, 2013). A non-parametric algorithm, not controlled by a set of statistical parameters, offers a more flexible learning process of separation for various types and bigger size data. Support vector machine (SVM) is the most popular non-parametric classifier due to its performance, ability to generalize problems and uniqueness in delivering global optimum solution (Awad and Khanna, 2015; Lu and Weng, 2007). Compared to the MLE algorithm, SVM has demonstrated its robustness in land use classification using TM, OLI, and Sentinel-2 images with SVM's to MLE's OA sequentially comparing as follows: 86.67 % to 80%, 81.67% to 70.60%, and 84.17% to 76.40% (Taati et al., 2015; Topaloğlu et al., 2016). However, SVM is outperformed by multi-layered machine learning algorithms such as classification trees and naive Bayes [22]. Nevertheless, no single algorithm can solve every LULC classification problem. De Fries & Chan (DeFries and Chan, 2000) compared multi-layered decision tree algorithms with bootstrap

DOI: 10.31315/ijgev.0203

aggregation (DTbagging) and boosting (DTboosting) to compare with the standard decision tree. They found that while DTboosting produced the highest accuracy, DTbagging produced the highest stability and robustness to noise. Further, Friedman (Friedman, 2002) established a mixed bagging and boosting model namely stochastic gradient boosting (SGB) to improve calculation speed and model performance which is successfully implemented by Lawrence (Lawrence et al., 2004) using two high-resolution images, IKONOS and Probe-1.

The selection of a machine learning algorithm for LULC classification is determined by the quality and availability of data in the area of interest (Mishra et al., 2017). However, retrieving and collecting information for supervised LULC classification is quite difficult, especially in cities in a developing country with a tropical climate due to scarcity of cloud-free spatial data and the spectral similarity among urban and non-urban features of land surface due to the mixed-land use in the city region (Schneider et al., 2015). Yogyakarta is a rapidly growing secondary city in Indonesia that is well-recognized for regional-ranked universities in Southeast Asia, such as the University of Gadjah Mada and the University of Islam Indonesia (UII). Since 1947, the local government of Yogyakarta city has divided its peri-urban into northern zone (for campus and academic-related development) and southern zone (for cultural-related development) (Yunus, 1991). From 1980 to 2000, Yogyakarta Special Province has experienced a significant increase in population as well as length of road and built up area, showing that Sleman Regency (the administrative area in northern zone) has the largest decrease in the agricultural area followed by Bantul Regency (the administrative area in southern zone) (Yunus and Harini, 2005). The city has been experiencing sprawl phenomenon to its peri-urban since the 2000s (Giyarsih, 2001; Rachmawati et al., 2004). Cellular automata (CA) model integrated with binary logistic regression/Markov chain using medium spatial-resolution images showed LULC change in Yogyakarta city, and its agglomeration area increased at a rate of 329 ha/year in 2003-2013 mainly to north-eastern region and predicted to increase to 539 ha/year in 2013-2023 mainly to south-western direction (Wijaya and Umam, 2015). However, the CA-Markov Chain model is a one-way statistical predictor that works based on contiguity rule and cannot calculate the fluctuation of the change trends (Memarian et al., 2012). Sustainable and long-term oriented land management as part of the principles of good governance (Gochero, 2018) cannot be put into practice unless a comprehensive knowledge of the LULC change map is maintained on a regular basis.

Periodical spatio-temporal assessment of urbanization is urgently needed to support various policies and decisions making in sustainable land resource management, not only for government in developing countries but also in small developed countries. Despite freely available remotely sensed data, technical challenges in data mining have hindered the updates of periodic land cover maps by the local government, especially in the less developed countries such as in the Lower Mekong and Hindu Kush-Himalaya (Saah et al., 2019) as well as in Libya (Alawamy et al., 2020). This has increased the need for single-date classifiers for land cover detection to support the capacity of local government in a self-sustained land use mapping to achieve the SDGs target in the future. Therefore, we tried to employ a supervised classifier model using a single medium spatial-resolution imagery data to produce a coherent multi-temporal land cover change detection based on empirical evidence from Yogyakarta city that represents a rising city in a developing country. The aims of this study are: (i) to establish a robust and reliable land cover classifier model with a machine learning algorithm using single medium spatial resolution; and (ii) to assess the replicability of the classifier model to multi-dates of medium spatial resolution.

2. Study Area

The target study area is 38 districts in Kartamantul, the urban agglomeration area in Yogyakarta Province, Indonesia, which consists of three adjacent municipalities as shown in Fig.1; Yogyakarta city

DOI: 10.31315/ijgev.0203

(14 districts), Sleman Regency (14 districts), and Bantul Regency (10 districts). Total area is 77,170 hectares and population number is 2,037,928 in 2019 (BPS-Statistics of Bantul Regency, 2019; BPS-Statistics of Sleman Regency, 2019; BPS-Statistics of Yogyakarta Municipality, 2019). Kartamantul area is predominantly covered by agricultural land (paddy field, cropland, and plantation) encircling the urban area (Yogyakarta city) in the middle. Yogyakarta city had been the capital of Indonesia from 1949 to 1950 and started to be developed into an academic-related activity area in the north (now Sleman Regency) and Mataram Kingdom culture-related activity area in the south (now Bantul Regency) (Yunus, 1991). In 2000, Yogyakarta Province had the highest urbanization level nationwide at 57.7% (Widyatmoko, 2007) whereas the urbanization level in peri-urban area reached 37.77% (Selang et al., 2018) in 2016. The study area is located in 7.33°S, 110.50°E and situated at an elevation from 45 m to 500 m above sea level (asl). Geologically, the province is home to an active volcano of >1000 m asl in the north and the Indian Ocean in the south of the area, making the city into a basin area. Moreover, the rainy season lasts for half a year (from November to April) with an average precipitation of 221.5 mm/month, an annual average temperature of 26.5°C and an average humidity of 85% has made the climate good for agriculture-based activity (BPS-Statistics of Bantul Regency, 2005; BPS-Statistics of Sleman Regency, 2005; BPS-Statistics of Yogyakarta Municipality, 2005).

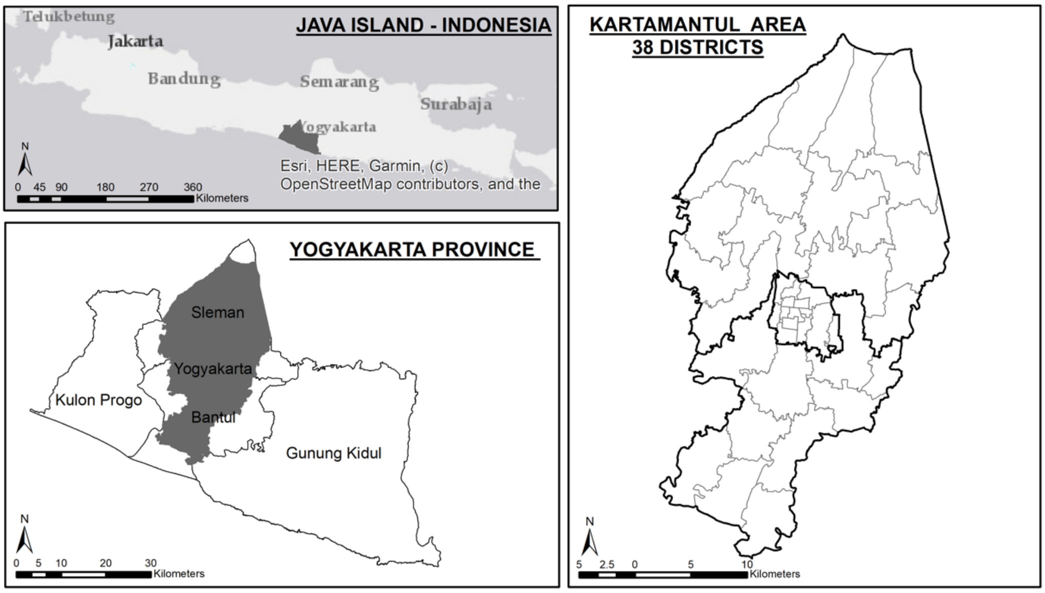

Fig. 1. Location of Kartamantul as a study area with 38 districts, consisting of 14 districts in Sleman Regency, 14 districts in Yogyakarta city, and 10 districts in Bantul regency of Yogyakarta Province, Java Island, Indonesia. Copyright: ESRI OpenStreetMap contributor.

References

ABP, 2020. 14 Tahun Tragedi Gempa Yogyakarta, Berikut Fakta-Faktanya [WWW Document]. https://news.okezone.com. URL https://news.okezone.com/read/2020/05/27/510/2220403/14-tahun-tragedi-gempa-yogyakarta-berikut-fakta-faktanya. (accessed 12.15.20).

Afaq, Y., Manocha, A., 2021. Analysis on change detection techniques for remote sensing applications: A review. Ecol. Inform. https://doi.org/10.1016/j.ecoinf.2021.101310

Alawamy, J.S., Balasundram, S.K., Hanif, A.H.M., Sung, C.T.B., 2020. Detecting and analyzing land use and land cover changes in the Region of Al-Jabal Al-Akhdar, Libya using time-series landsat data from 1985 to 2017. Sustain. 12. https://doi.org/10.3390/su12114490

Albon, C., 2017. Nested Cross Validation [WWW Document]. chrisalbon.com. URL https://chrisalbon.com/machine_learning/model_evaluation/nested_cross_validation/ (accessed 5.14.18).

Awad, M., Khanna, R., 2015. Efficient learning machines: Theories, concepts, and applications for engineers and system designers, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. https://doi.org/10.1007/978-1-4302-5990-9

Bharly, Y., 2006. Pengembangan Terminal Penumpang Bandar Udara Adisutjipto Sebagai Bandara Internasional. Universitas Diponegoro.

BPS-Statistics of Bantul Regency, 2019. Kabupaten Bantul Dalam Angka 2019. BPS-Statistics of Bantul Regency, Bantul.

BPS-Statistics of Bantul Regency, 2005. Kabupaten Bantul Dalam Angka 2005. BPS-Statistics of Bantul Regency, Bantul.

BPS-Statistics of Sleman Regency, 2019. Kabupaten Sleman Dalam Angka 2019. BPS-Statistics of Sleman Regency, Sleman.

BPS-Statistics of Sleman Regency, 2005. Kabupaten Sleman Dalam Angka 2005. BPS-Statistics of Sleman Regency, Sleman.

BPS-Statistics of Yogyakarta Municipality, 2019. Kota Yogyakarta Dalam Angka 2019. BPS-Statistics of Yogyakarta Municipality, Yogyakarta.

BPS-Statistics of Yogyakarta Municipality, 2005. Kota Yogyakarta Dalam Angka 2005. BPS-Statistics of Yogyakarta Municipality, Yogyakarta.

Brownlee, J., 2020. Nested Cross-Validation for Machine Learning with Python [WWW Document]. Mach. Learn. Mastery Pty. Ltd. URL https://machinelearningmastery.com/nested-cross-validation-for-machine-learning-with-python/ (accessed 12.1.20).

Brownlee, J., 2014. Classification Accuracy is Not Enough: More Performance Measures You Can Use [WWW Document]. Mach. Learn. Mastery Pty. Ltd. URL https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

Cochrane, C., 2018. Time Series Nested Cross-Validation [WWW Document]. Towar. Data Sci. URL https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9 (accessed 2.18.20).

Cohen, B., 2004. Urban growth in developing countries: A review of current trends and a caution regarding existing forecasts. World Dev. 32, 23–51. https://doi.org/10.1016/j.worlddev.2003.04.008

DeFries, R.S., Chan, J.C.-W., 2000. Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data. Remote Sens. Environ. 74, 503–515.

Folta, W., 2017. Machine Learning Tip: Nested Cross Validation – When (Simple) Cross Validation Isn’t Enough [WWW Document]. Mach. Learn. Times. URL http://www.predictiveanalyticsworld.com/machinelearningtimes/nested-cross-validation-simple-cross-validation-isnt-enough/8952/ (accessed 5.14.18).

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Geneletti, D., La Rosa, D., Spyra, M., Cortinovis, C., 2017. A review of approaches and challenges for sustainable planning in urban peripheries. Landsc. Urban Plan. 165, 231–243. https://doi.org/10.1016/j.landurbplan.2017.01.013

Giyarsih, S.R., 2001. Gejala Urban Sprawl Sebagai Pemicu Proses Densifikasi Permukiman di Daerah Pinggiran Kota (Urban Fringe Area) Kasus Pinggiran Kota Yogyakarta. J. Perenc. Wil. dan Kota 12, 40–45.

Gochero, P., 2018. Econometric Analysis of Foreign Direct Investment in the Zimbabwean Mining Sector 2005-2014. Theor. Econ. Lett. 08, 3157–3177. https://doi.org/10.4236/tel.2018.814196

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS J. Photogramm. Remote Sens. 116, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

GT, NIQ, ESP, 2018. Peduli Lingkungan, Mahasiswa UII Ikuti Kegiatan Tanam Pohon [WWW Document]. www.uii.ac.id. URL https://www.uii.ac.id/peduli-lingkungan-mahasiswa-uii-ikuti-kegiatan-tanam-pohon/%0D%0A%0D%0A (accessed 12.13.20).

Gutierrez, D.D., 2015. Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R, First Edit. ed. Technics Publications, New Jersey.

Hansen, C., 2019. Nested Cross-Validation Python Code [WWW Document]. Mach. Learn. From Scratch. URL https://mlfromscratch.com/nested-cross-validation-python-code/#/

Heydari, S.S., Mountrakis, G., 2018. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. Remote Sens. Environ. 204, 648–658. https://doi.org/10.1016/j.rse.2017.09.035

Huriati, N., 2008. Urban Fringe Area Development in Yogyakarta City 1992-2006. Universitas Indonesia.

Kaliraj, S., Chandrasekar, N., Ramachandran, K.K., Srinivas, Y., Saravanan, S., 2017. Coastal landuse and land cover change and transformations of Kanyakumari coast, India using remote sensing and GIS. Egypt. J. Remote Sens. Sp. Sci. 20, 169–185. https://doi.org/10.1016/j.ejrs.2017.04.003

Kodoatie, R.J., 2012. Tata Ruang Air Tanah, 1st ed. Andi Offset, Yogyakarta.

Kumar, Y., Sahoo, G., 2012. Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA. Int. J. Inf. Technol. Comput. Sci. 7, 43–49. https://doi.org/10.5815/ijitcs.2012.07.06

Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sens. Environ. 90, 331–336. https://doi.org/10.1016/j.rse.2004.01.007

Li, L., Wu, Y., Ye, M., 2014. Multi-class image classification based on fast stochastic gradient boosting. Inform. 38, 145–153.

Li, W., Liu, Z., 2011. A method of SVM with normalization in intrusion detection. Procedia Environ. Sci. 11, 256–262. https://doi.org/10.1016/j.proenv.2011.12.040

Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2015. Remote Sensing and Image Interpretation, Seventh. ed. John Wiley & Sons, Inc.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870. https://doi.org/10.1080/01431160600746456

Memarian, H., Balasundram, S.K., Sung, C.T.B., Talib, J. Bin, Sood, A.M., Abbaspour, K., 2012. Validation of CA-Markov for Simulation of Land Use and Cover Change in the Langat Basin , Malaysia. J. Geogr. Inf. Syst. 4, 542–554. https://doi.org/10.4236/jgis.2012.46059

Mishra, S., Shrivastava, P., Dhurvey, P., 2017. Change Detection Techniques in Remote Sensing: A Review. Int. J. Wirel. Mob. Commun. Ind. Syst. 4, 1–8. https://doi.org/10.21742/ijwmcis.2017.4.1.01

Mohajane, M., Essahlaoui, A., Oudija, F., Hafyani, M. El, Hmaidi, A. El, Ouali, A. El, Randazzo, G., Toedoro, A.C., 2018. Land Use/Land Cover (LULC) Using Landsat Data Series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco. Environments 5, 1–16. https://doi.org/10.3390/environments5120131

Najamudin, I., 2017. Flight Safety Case Study : Adi Sucipto Airport Jogjakarta-Indonesia. Int. Ref. J. Eng. Sci. 6, 29. https://doi.org/10.183x/C6811932

OptunaContributors, 2018. Welcome to Optuna's documentation! [WWW Document]. https://github.com/. URL https://optuna.readthedocs.io/en/v0.19.0/

Phyton, 2019. Python 3.6.10 [WWW Document]. https://www.python.org/. URL https://www.python.org/downloads/release/python-3610/

Provinsi Jawa Tengah, D.P.M.D.P.T.S.P., 2019. Bandara Internasional Adi Sucipto Yogyakarta [WWW Document]. https://web.dpmptsp.jatengprov.go.id. URL https://web.dpmptsp.jatengprov.go.id/sarpras/2/28 (accessed 12.15.20).

Rachmawati, R., Rijanta, R., Subanu, L.P., 2004. Peranan Kampus Sebagai emicu Urbanisasi Spasial di Pinggiran Kota Yogyakarta. Maj. Geogr. Indones.

Radočaj, D., Obhođaš, J., Jurišić, M., Gašparović, M., 2020. Global open data remote sensing satellite missions for land monitoring and conservation: A review. Land 9, 1–24. https://doi.org/10.3390/land9110402

Richards, J.A., 2013. Remote Sensing Digital Image Analysis, An Introduction, Fifth. ed. Springer, Berlin. https://doi.org/10.1007/978-3-642-30062-2

Rocchini, D., Petras, V., Petrasova, A., Horning, N., Furtkevicova, L., Neteler, M., Leutner, B., Wegmann, M., 2017. Open data and open source for remote sensing training in ecology. Ecol. Inform. 40, 57–61. https://doi.org/0.1016/j.ecoinf.2017.05.004

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., Nguyen, Q.H., Patterson, M., Johnson, G., Markert, K., Flores, A., Anderson, E., Weigel, A., Ellenberg, W.L., Bhargava, R., Aekakkararungroj, A., Bhandari, B., Khanal, N., Housman, I.W., Potapov, P., Tyukavina, A., Maus, P., Ganz, D., Clinton, N., Chishtie, F., 2019. Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. Front. Environ. Sci. 7. https://doi.org/10.3389/fenvs.2019.00150

Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10. https://doi.org/10.1371/journal.pone.0118432

Salari, N., Shohaimi, S., Najafi, F., Nallappan, M., Karishnarajah, I., 2014. A novel hybrid classification model of genetic algorithms, modified k-nearest neighbor and developed backpropagation neural network. PLoS One 9. https://doi.org/10.1371/journal.pone.0112987

Sandler, A.M., Rashford, B.S., 2018. Misclassification error in satellite imagery data: Implications for empirical land-use models. Land use policy 75, 530–537. https://doi.org/10.1016/j.landusepol.2018.04.008

Schneider, A., Mertes, C.M., Tatem, A.J., Tan, B., Sulla-Menashe, D., Graves, S.J., Patel, N.N., Horton, J.A., Gaughan, A.E., Rollo, J.T., Schelly, I.H., Stevens, F.R., Dastur, A., 2015. A new urban landscape in East-Southeast Asia, 2000-2010. Environ. Res. Lett. 10. https://doi.org/10.1088/1748-9326/10/3/034002

Selang, M.A., Iskandar, D.A., Widodo, R., 2018. Tingkat Perkembangan Urbanisasi Spasial Di Pinggiran KPY (Kawasan Perkotaan Yogyakarta) Tahun 2012-2016. Kota Layak Huni "Urbanisasi dan Pengemb. Perkota. 32–40.

Taati, A., Sarmadian, F., Mousavi, A., Pour, C.T.H., Shahir, A.H.E., 2015. Land use classification using support vector machine and maximum likelihood algorithms by landsat 5 TM images. Walailak J. Sci. Technol. 12, 681–687. https://doi.org/10.14456/vol12iss11pp

Tewabe, D., Fentahun, T., 2020. Assessing land use and land cover change detection using remote sensing in the Lake Tana Basin, Northwest Ethiopia. Cogent Environ. Sci. 6, 1778998. https://doi.org/10.1080/23311843.2020.1778998

Topaloğlu, R.H., Sertel, E., Musaoğlu, N., 2016. Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 41, 1055–1059. https://doi.org/10.5194/isprsarchives-XLI-B8-1055-2016

U.S. Geological Survey, n.d. Earth Explorer [WWW Document]. https://earthexplorer.usgs.gov/. URL https://earthexplorer.usgs.gov/

UKEssays, 2018. Supervised Image Classification Techniques [WWW Document]. URL https://www.ukessays.com/essays/engineering/supervised-image-classification-9746.php?vref=1 (accessed 3.23.21).

UN, 2020. THE 17 GOALS | Sustainable Development [WWW Document]. Dep. Econ. Soc. Aff. URL https://sdgs.un.org/goals

USGS National Land Imaging Program, 2017. Landsat Collection 1 Level 1 26.

Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7. https://doi.org/10.1186/1471-2105-7-91

Widyatmoko, M.R., 2007. Proses Urbanisasi Perdesaan di Daerah Istimewa Yogyakarta dan Urbanisasi di Indonesia yang Melatarbelakanginya. University of Gadjah Mada.

Wijaya, M.S., Umam, N., 2015. Pemodelan Spasial Perkembangan Fisik Perkotaan Yogyakarta Menggunakan model Cellular Automata dan Regresi Logistik Biner. Maj. Ilm. Globe Vol. 17 No.2 Desember 2015 165–172.

XGBoostDevelopers, 2020. Get Started with XGBoost [WWW Document]. https://xgboost.readthedocs.io/. URL https://xgboost.readthedocs.io/en/latest/get_started.html

Yang, W., Xu, L., Chen, X., Zheng, F., Liu, Y., 2015. Chi-squared distance metric learning for histogram data. Math. Probl. Eng. https://doi.org/10.1155/2015/352849

Yunus, H.S., 1991. The Evolving Urban Planning: The Case of The City of Yogyakarta. Indones. J. Geogr. 21, 1–14. https://doi.org/10.22146/ijg.2193

Yunus, H.S., Harini, R., 2005. The dominant factors affecting agricultural land use (rice field) change in Yogyakarta Special Province 37. https://doi.org/10.22146/ijg.2217

Zou, Q., Xie, S., Lin, Z., Wu, M., Ju, Y., 2016. Finding the Best Classification Threshold in Imbalanced Classification. Big Data Res. 5, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001

Appendix A

Notably, there were two spots, such as Adisucipto International Airport (AIA) in the mid-eastern part (in a blue-lined square) and University of Islam Indonesia (UII) in the northern part (in a red-lined square), which had been perfectly classified as class-U in 1999 and 2005 were then classified as class-NU in 2011 (see Fig.8). This was most probably due to the changes in vegetational cover conditions in 2011. Apart from time to time comparison of satellite images for those two spots presented in Fig. A1 below (e.g., from 2006 to 2011 for UII and from 2006 to 2010 for AIA), information from newspaper articles and an in-depth interview with local inhabitants were collected to confirm the phenomenon.

DOI: 10.31315/ijgev.0203

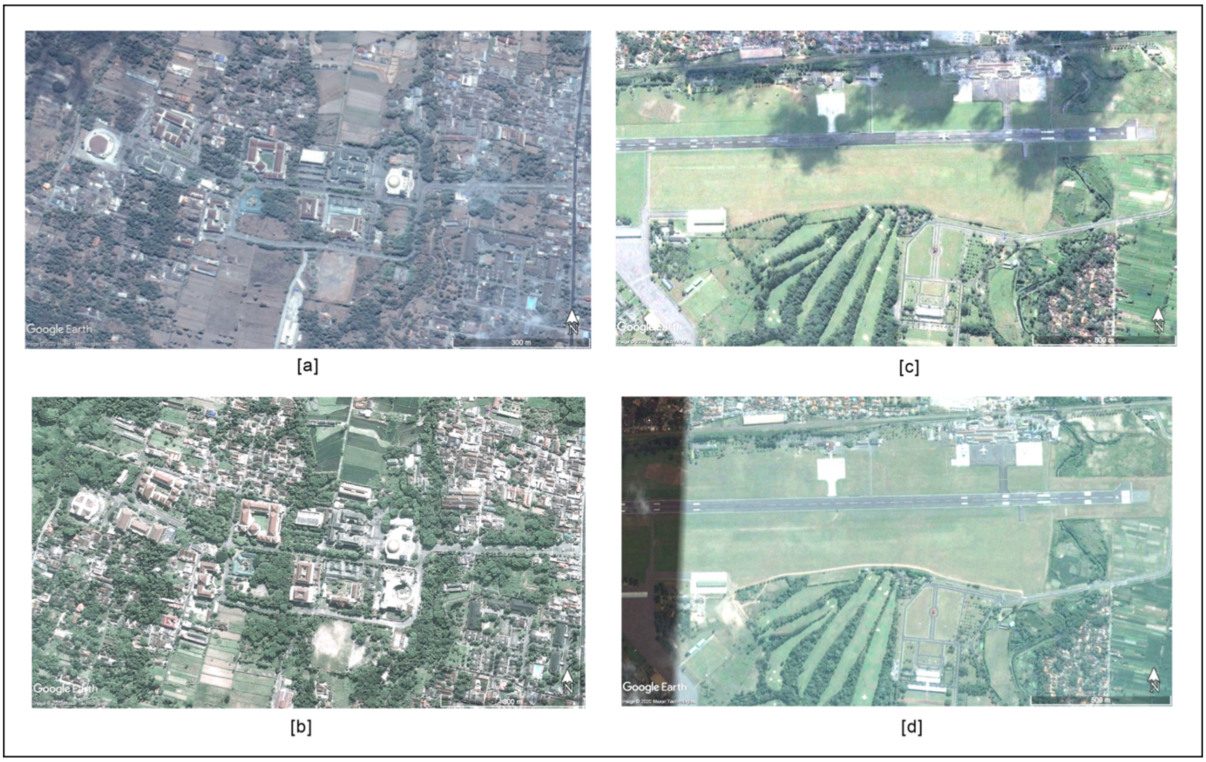

Fig. A1. Google Earth Historical Imagery to corroborate the classification result in Yogyakarta city from 2005 to 2011 featuring (a) UII in Sept-2006. Source: Google Earth pro V 7.3. (September 11, 2006). University of Islam Indonesia, Yogyakarta Indonesia 7° 41' 17.27"S, 110° 24' 51.09"E, Eye alt 827 m.Maxar Technologies 2020. http://www.earth.google.com [August 15, 2020]; (b) UII in Jun-2011. Source: Google Earth pro V 7.3. (June 11, 2011). University of Islam Indonesia, Yogyakarta Indonesia 7° 41' 17.27"S, 110° 24' 51.09"E, Eye alt 827 m.Maxar Technologies 2020. http://www.earth.google.com [August 15, 2020]; (c) AIA in Jun-2006. Source: Google Earth pro V 7.3. (June 8, 2006). Adisucipto International Airport, Yogyakarta Indonesia 7° 47' 28.39"S, 110° 25' 52.33"E, Eye alt 4480 m.Maxar Technologies 2020. http://www.earth.google.com [August 15, 2020]; (d) AIA in Apr-2010. Source: Google Earth pro V 7.3. (April 8, 2010). Adisucipto International Airport, Yogyakarta Indonesia 7° 47' 28.39"S, 110° 25' 52.33"E, Eye alt 4480 m.Maxar Technologies 2020. http://www.earth.google.com [August 15, 2020].

UII campus was built in the 1990s, covering almost 39 hectares of water recharge area in Ngemplak District, Sleman Regency. In 1995, UII initiated to create an environmentally friendly campus by planting more vegetations on its surrounding according to a special zoning plan (see Fig.A2), such as bamboo along the riverbanks within the campus, indigenous trees from the west, middle and east part of Indonesia, symbolic trees of Java Island such as *Manilkara kauki* and *Stelechocarpus burahol* (GT et al., 2018). Those trees took years to grow that it was not quite shown in 2005, but it was formidably shown in 2011 where the foliage was already very thick and dense.

DOI: 10.31315/ijgev.0203

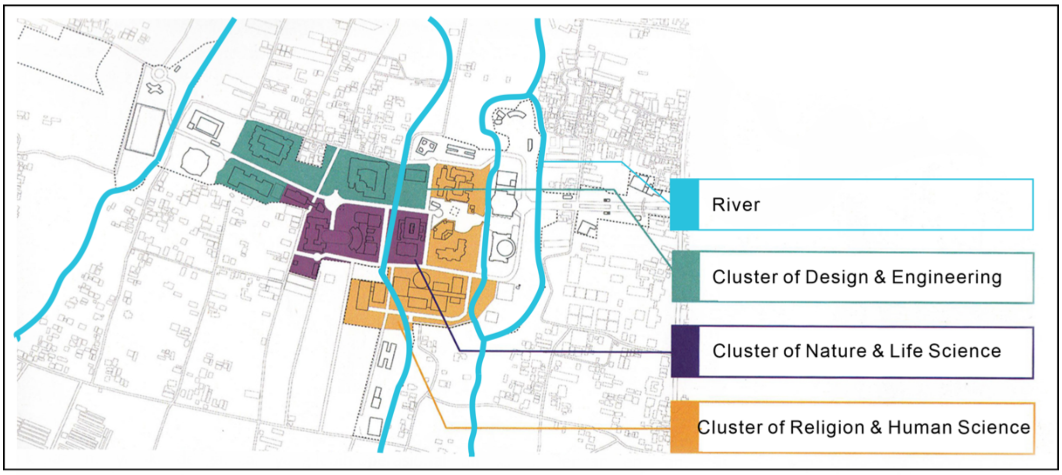

Fig.A2. Spatial Zoning Plan for Greening the UII Campus (modified from (GT et al., 2018)).

Meanwhile, the case of AIA, with a total area of about 180 hectares (Provinsi Jawa Tengah, 2019), has a runway of 2200 m x 45 m that was just coated in 2005 (Najamudin, 2017) so that it could be reflected perfectly in 2005 imagery. The 2006 Yogyakarta earthquake catastrophe cracked in the AIA runway and ruined the AIA terminal building so badly that the airport was closed down for about 10 days (ABP, 2020; Bharly, 2006). The reconstruction of the runway was probably not done seamlessly because of the rush to start the operation for transportation gate of volunteers and donations from all over the world for the earthquake recovery. As a result, the runway was not as clearly reflected in 2011 as it was in 2005. Furthermore, it appeared that the grass that was supposed to cover the 2285 m x 150 m solid soil based-runway strip (Najamudin, 2017) took years to thrive, so it was reflected pungently in 2011 rather than in 2006. Fig. A3 presents a detail of the runway map of AIA.

Fig. A3. Map of Runway in Adisutjipto International Airport -AIA (taken from (Najamudin, 2017)).