



Implementing Random Forest Algorithm in GEE: Separation and Transferability on Built-Up Area in Central Java, Indonesia

Aninda W. Rudiastuti^{a,*}, Yustisi Lumban-Gaol^a, Florence E. S. Silalahi^a, Yosef Prihanto^a,
Widodo S. Pranowo^{b,c}

^a National Research and Innovation Agency of Indonesia, Cibinong, 16911, Indonesia

^b Marine and Coastal Data Laboratory, Research & Development Center for Marine & Coastal Resources, North Jakarta, 14430, Indonesia

^c Department of Hydrography, Indonesian Naval Postgraduate School (STAL), Jakarta

Corresponding author: *yustisiardhitasari@gmail.com

Abstract— Measuring the status of achievement of the SDGs is the task and concern of many countries in the world, including Indonesia. Indicators for achieving the SDGs enclose three main pillars, namely environmental, economic, and social. The updated land use/land cover information is needed for environmental pillars. One imperative land cover information is built-up land, which acts as a detector for expanding urban areas and measuring SDGs' target indicators. Indonesia's cultural diversity affects the distribution pattern of built-up land, especially settlements. This is a challenge in the up-to-date and rapid mapping of built-up land. This research aims to analyze the ability and transferability of the Random Forest model for built-up areas and settlements using Google Earth Engine (GEE) in Banyumas, Cilacap, and Tegal. Around 19 predictors from multi-sources satellites are integrated to identify four land cover classes. Discussion on predictor composition to improve model accuracy also carried on. The results showed that the algorithm separated four land cover classes, with the highest accuracy for separating water bodies and other classes (vegetation and open land), OA above 90%. Machine confusion regarding the separation between housing classes and other buildings was still found (F1 score 0.67 - 0.69). Applying the model to the other two areas resulted in a similar statistical trend to the trained model. However, the classification method developed in this paper can assist in the rapid description of land cover if up-to-date data from official sources are not available.

Keywords— Random forest; machine learning; Google Earth Engine (GEE); settlement; sentinel; Land Use/Land Cover (LULC).

Manuscript received 3 Jan. 2022; revised 25 Feb. 2022; accepted 4 Mar. 2022. Date of publication 31 Mar. 2022.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Indonesia has ratified the Sustainable Development Goals (SDGs) that have been set by the United Nations (UN) and the World Bank. The SDGs aim for countries to achieve the essential quality of life by 2030. One of the essential elements of sustainable development is conserving and managing natural resources. Thus, all aspects related to resources in an area are essential to be translated spatially so that the achievement of the SDGs can be monitored, and the status of sustainable development achievements can be reported regularly [1]. Several indicators highly depend on land cover data availability, focusing on the environmental pillar of SDGs Indonesia. For mention, indicator 11.1.1. (a) The percentage of households with access to adequate and affordable housing, indicator 11.2.1. (a) The proportion of the population with convenient access to public transportation, indicator 11.3.1. (a) The ratio of the expansion rate of

developed land to the population growth rate, indicator 11.6.1. (a) The percentage of households in urban areas whose waste management is served, indicator 11.7.1. (a) The proportion of urban open space for all, and Indicator 15.1.1* proportion forest area to the total land area.

Some even require data on the distribution of population/households and built-up land, which are projections of the center of human distribution. Thus, the urgency of providing and updating maps of built-up land, especially settlements, cannot be avoided. Land use (LU) and land cover (LC) are two types of cartographic information that are used in conjunction with one another for urban planning and environmental monitoring [2].

Various methods have been attempted to provide LULC maps using remote sensing data, especially on the built-up area mapping. Starting from the visual interpretation technique [3]–[7], pixel-based [8]–[10] and object-based classification [11]–[13], to artificial intelligence [14], [15]. However, no standard method has been established for

compiling the map. In a decade, the development of artificial intelligence (AI) computing technology, machine learning (ML), has been in great demand, especially for land cover mapping [16]–[20]. Artificial intelligence opens opportunities for its use in accelerating land cover mapping. Machine Learning (ML) techniques are widely used in spatial data and have the characteristics of big data, which often involves the process of earth feature extraction algorithms in the analysis process.

ML techniques have been used for a long time in the geospatial field. ML illustrates how we can detect object detection [21], such as cars, trains, and others. It is not an old technique, but its content is not far from conventional statistics like binary. There is some data that we take advantage of with high-resolution satellite imagery. Rousset et al. [2] successfully enhanced the overall accuracy when adding the LC classification output of the dedicated deep learning architecture to the raw channels input of the deep learning LU classification task (from 51% to 63% of overall accuracy). Meanwhile, Fagua and Ramsey [22] successfully developed annual LULC maps in the Chocó-Darien Global Ecoregion from 2002 to 2015 (kappa index 0.87) from a combination of the Random Forest ensemble learning classification tree system on cloud-free MODIS vegetation index products. Tsai et al. [19] stated that using a cloud-based Google Earth Engine platform to act as an advanced machine learning image classification should implement tuning techniques to find optimal classifier parameters (for example, the number of trees and the number of features for Random Forest classifier).

In addition to answering the SDGs targets, the availability of rapid land cover mapping, especially for built-up land and settlements, is practically beneficial for activities in the social sector. Among others are household poverty [23] and disaster management [24]–[26].

Indonesia's cultural diversity influences the distribution pattern of built-up land, especially settlements. This condition creates a challenge in itself in built-up mapping land that is up to date and can be done quickly. Beforehand, Rudiastuti et al. [27] carried out a brief study comparing various supervised learning algorithms in Google Earth Engine based on its accuracy for land use mapping in Purwokerto within an area of 81km². In this research, by modifying the test sample and expanding the area to accommodate more diversity, we strived to create a land cover model comprised of four land cover types: vegetation, built-up land (divided into non-residential and residential building classes), and water body. Banyumas Regency was chosen as the initial location for model development because Banyumas has both urban and rural areas. Subsequently is the possibility of the transferability model that has been built in Banyumas to map buildings and settlements in other areas, in this case, the Cilacap Regency and Tegal Regency.

The idea of separating buildings and settlements comes from two points. The first is the need for detailed settlement data as a challenge to answer the SDGs targets. The second is to implement Ji et al. method [28], which has succeeded in detecting settlements in rural areas using sentinel. The implementation of this method also looks at the performance of the supervised learning algorithm at GEE in carrying out research objectives. Several assumptions were applied in this

study, namely (1) sample selection considering the function of the building and the appearance of the building (between non-residential and residential buildings) (2) building size 100 m², so it is assumed that one-pixel Sentinel-2 represents one house. This paper also discusses the predictor composition that may increase model accuracy.

II. MATERIAL AND METHOD.

A. Study Area

The research took place at three locations in the Central Java province, namely Cilacap, Banyumas, and Tegal district. Those districts are sequentially located in Central Java Province's region's south, middle, and north (Figure 1). They differ in terms of geographic condition, where Cilacap and Tegal have coastal areas while Banyumas does not. We choose areas in the same province to avoid the predominant cultural diversity in building settlements.

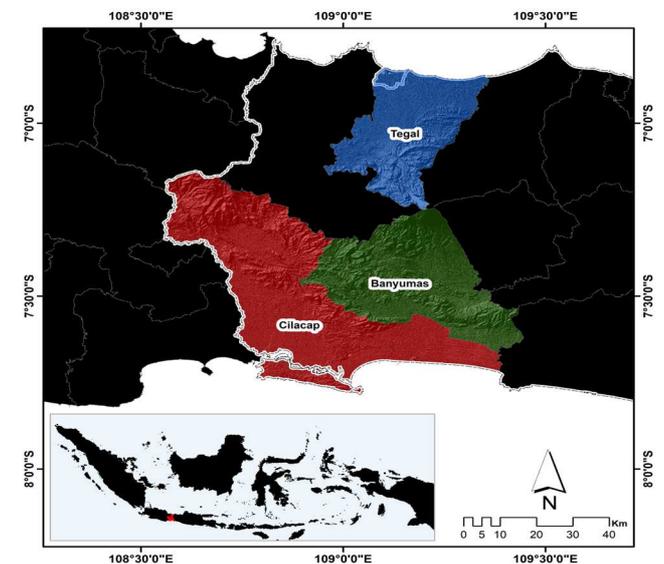


Fig. 1 Three study areas in Central Java. The bold white line is the borderline of Central Java Province. The blue, green, and red areas are Tegal, Banyumas, and Cilacap districts.

B. Data

This study uses six data types to define the model predictors: pure spectral information from the visual and NIR channel, the image transformation through several indexes, surface texture, night-time light, slope, and land surface temperature (LST). They were acquired from several data sources that are publicly open. Table 1 describes the required dataset in detail, including sources and specifications. Some were collected or computed using remote sensing data, from both active and passive sensors, with various spatial and temporal resolutions. Images from Sentinel-1 (SAR), Sentinel-2, Landsat-8, and Suomi NPP VIIRS within the time range from 1 June 2020 to 31 May 2021 became data sources for most model predictors.

Surface reflectance characteristics were captured using Sentinel-2 Level 2A, whereas synthetic aperture radar (SAR) images provide information on surface materials' structure and dielectric properties using radar imagery. Furthermore, night-time light (NTL) imagery identifies the level of human activity and, as a result, can significantly provide antecedent

probabilities of impervious surfaces occurrence in a given area [29].

According to Kafy et al. [30], the changes in urban areas and raising LST have a strong correlation. Thus, we add the LST predictor assuming it will affect the machine's ability to differentiate between each land cover class, especially built-

up land. Besides the model predictors data, we collect the administrative boundary from the National Base Map to distinguish each study area. In addition, the National Digital Elevation Model of Indonesia (DEMNAS) is used to generate slope.

TABLE I
DATA

Data	Source	Resolution/Specification
Spectral reflectance:		
1. Red band		
2. Green band		
3. Blue band		
4. NIR band		
Spectral indices:		
5. Normalized Difference Built-Up Index [21], [32]–[36]		
$NDBI = \frac{B_{11} - B_8}{B_{11} + B_8}$		
6. New Built-Up Index [36], [37]		
$NBI = \frac{B_4 * B_{11}}{B_8}$		<ul style="list-style-type: none"> • resolution varies from 10, 20, 60 meters (depend on bands)
7. Built-up or Bareness Index [28]		<ul style="list-style-type: none"> • Surface Reflectance
$BOBI = \frac{B_8 - B_4}{B_8 + B_4}$	Sentinel-2 (Multi-Spectral Instrument/MSI) Level 2a	<ul style="list-style-type: none"> • B₂ = Blue • B₃ = Green • B₄ = Red • B₈ = NIR • B₁₁ = SWIR
8. Ratio Resident-area Index [28], [34], [35]		
$RRI = \frac{B_2}{B_8}$		
9. Modified Normalized Difference Water Index [32], [33], [37]		
$MNDWI = \frac{B_3 - B_{11}}{B_3 + B_{11}}$		
10. Normalized Difference Vegetation Index [21], [32], [38], [39]]		
$NDVI = \frac{B_8 - B_4}{B_8 + B_4}$		
11. Soil salinity index (SI) [1], [11]		
$SI = \sqrt{B_2 * B_4}$		
12. Texture features: VV polarization	Sentinel 1 (Synthetic Aperture Radar/SAR)	<ul style="list-style-type: none"> • 10m
13. Texture features: VH polarization		<ul style="list-style-type: none"> • Wide swath (IW) acquisition mode
14. Texture features: GLCM_VH		
15. Texture features: GLCM_VV		<ul style="list-style-type: none"> • VV+VH polarization
16. Land surface temperature (LST)	Landsat 8 Tier 2	30m
17. Nighttime light (NTL) information	Suomi NPP VIIRS	463.83m
18. VANUI [28], [40]		
$VANUI = (1 - NDVI) * NTL$		
19. Slope	Seamless National Digital Elevation Model (DEM)	0.27-arcsecond \approx 8.25m
Administration boundary & Land Cover information	National Base Map (RBI) of Banyumas, Cilacap, and Tegal regency	1:25000

C. Method

In general, the workflow of this research consists of four main steps: data acquisition, data preparation for training, accuracy assessment of the model, and transfer learning analysis. Each process is executed using a cloud-based

geospatial processing platform, namely Google Earth Engine (GEE), except for the collection of labeled points. GEE comprehends a group of publicly available datasets that employ Google computer power, easily accessible since it is open to the public [31]. Figure 2 shows the flowchart of this study.

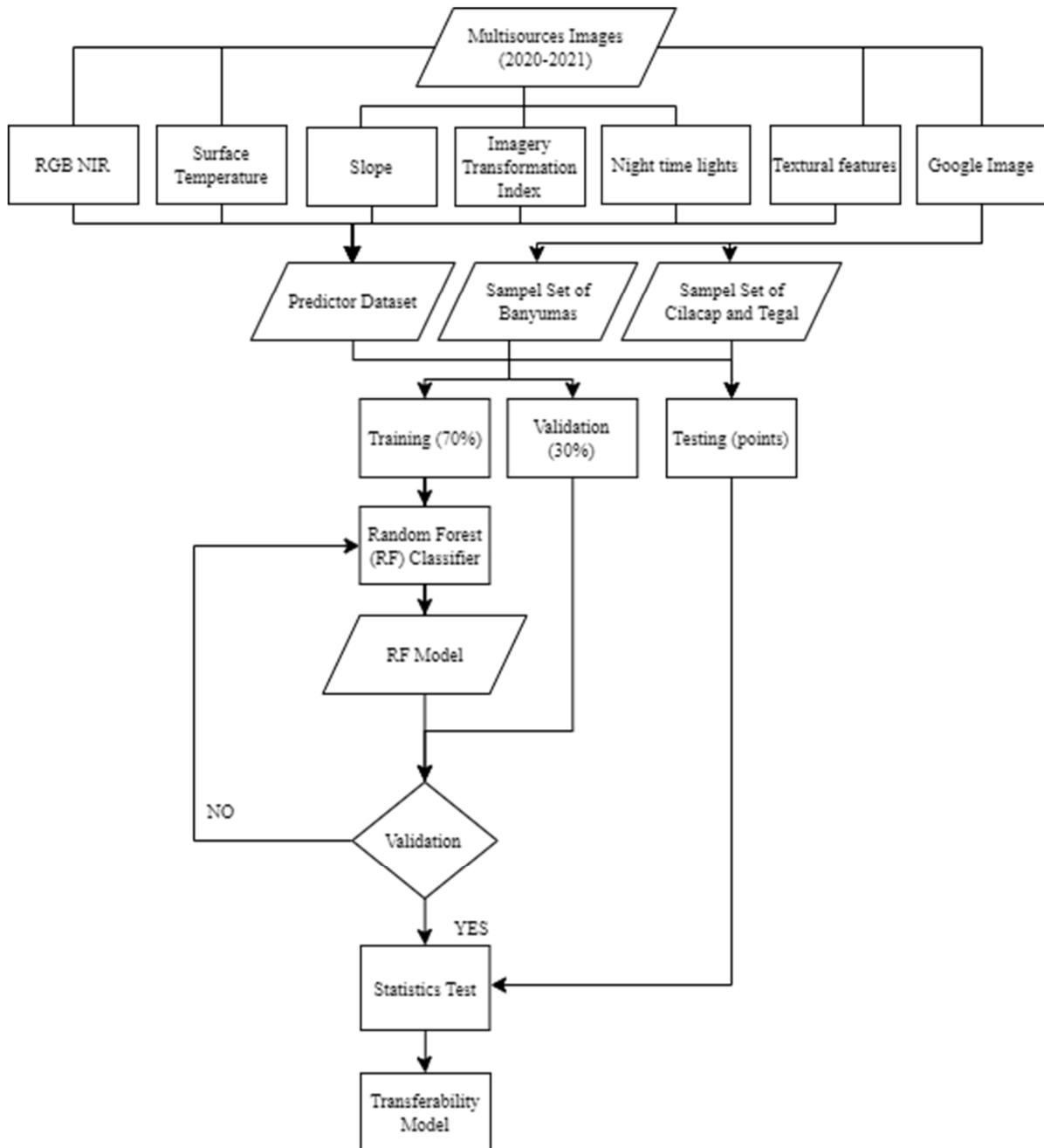


Fig. 2 Research workflow.

We manually label different objects into four classes using the Google Earth Pro platform [8]. Those classes are others, built-up, the settlements, and water bodies. The built-up class includes all non-residential structures such as office buildings, schools, roads, bridges, and industry and manufacturing, while the settlement class comprises multifunctional houses or residential buildings. All water objects, such as rivers and lakes, belong to the class of water bodies. Meanwhile, the unclassified category, or others, consists of the other three classes, such as bare soil, vegetation, and agricultural land. We only consider four classes since this research focuses on the capability of the model to detect and distinguish settlements from other built-up lands. These labeled points are collected in each area of interest (Table 2). There are 992 points in Banyumas, 416 in Cilacap, and 120 in Tegal. The

composition of the number of points for independent testing and modeling adjusts the total area of the study area.

TABLE II
CLASS CODE OF THE STUDY AREA

Area	Class Code			
	1 (others)	2 (Built-up)	3 (Settlement)	4 (Waterbody)
Banyumas	414	253	262	56
Cilacap	123	112	125	52
Tegal	34	34	39	13

The land cover classification is carried out using Random Forest (RF) algorithm, considering that this approach has better performance for detecting land cover types between built-up land and other classes in an area of 81 km² compared to other methods, such as Support Vector Machine and

Minimum Distance [27]. We used default parameters for the RF classifier, but we changed the numberOfTrees from 100 to 125 and set the variablesPerSplit to four. The training model computation is performed using the Banyumas dataset in this research. Therefore, labeled points of Banyumas are separated into training and testing sets using a 70:30 ratio of random sampling method, where 70% of points are used for training, and the rest is for testing. The sampling point is determined by the seed (pseudorandom number generator) [41].

A minimum of 10 iterations is applied using different seeds. The model with the highest overall accuracy will be recorded in this paper. The trained model is then reused in new locations, Cilacap and Tegal, to assess the model transferability. We try two combinations of predictors for the experiment. First, we used without LST (C1), and the other one is using LST (C2) to determine how machine learning detects four land cover classes and separates building types only by recognizing impervious surfaces (residential and non-residential).

Subsequently, the RF variable importance is computed after the training process to examine the contribution of each predictor during training. The variable of importance determines whether our predictors have equal contribution or are less important than others.

Finally, the pixel resolution of the classification output is determined to be 10m based on following the highest resolution of the predictor variable used (the Sentinel-2 resolution). Table 1 lists the different pixel resolutions of each image source used in research for the record. This follows the best available open-source image required in this research.

D. Accuracy Assessment

The predicted class generated using the RF algorithm was compared with the labeled points in the testing dataset employing a confusion matrix to evaluate the classification results. Here we use standard approaches to measure the performance of our model by calculating the overall accuracy (OA), prediction, recall, and F1 score. Mathematically, those metrics are written as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Where TP is the total number of true positives (e.g., settlement point is predicted as settlement point), TN is the total number of true negatives (e.g., the non-settlement point is predicted as a non-settlement point). FP is the total number of false positives (e.g., the non-settlement point is predicted as a settlement point). Meanwhile, FN is the number of false negatives (e.g., the settlement point is predicted as a non-settlement point).

The OA is a ratio that describes the number of correctly predicted points divided by the total number of points. The precision value represents a positive prediction: how many predicted settlement points are true settlement points. The recall value indicates the sensitivity or true positive value, the

number of true settlement points correctly predicted as settlement points. When there is no misclassification, a good prediction result should have a perfect precision and recall value of one or 100%. The F1 score showed the harmonic mean of precision and recall number of the model. It is calculated as the weighted average of precision and recall. As a result, this score considers both false positives and false negatives. It is not as intuitive as accuracy, but the F1 score is frequently more advantageous than accuracy, especially if the class distribution is unequal. Accuracy works best when the cost of false positives and false negatives is comparable. If the cost of false positives and false negatives is considerably different, it is best to consider precision and recall.

We compute OA for all classes in this study. Meanwhile, because this work focuses on settlement point classification, we calculate precision, recall, and F1 Scores like binary classification using settlement and non-settlement classes. We also compute the kappa coefficient using the confusion matrix to eliminate the evaluation bias due to correct classification by chance.

III. RESULT AND DISCUSSION

The statistics of the two predictor combinations are shown in Table 3. The overall accuracy for the classification model without accommodating information from the Thermal Infrared Sensor (C1) reaches 0.806, with a kappa coefficient of 0.717. The addition of LST variables (C2) gives different results to the classification model. By including all the predictors (19 variables in Table 1), the statistical information on the model in the results of C2 improves with the overall accuracy and kappa coefficient of 0.843 and 0.772, respectively.

TABLE III
THE ACCURACY OF THE CLASSIFICATION MODEL AT BANYUMAS REGENCY

Classification number	Validation				F1 score
	OA	Kappa coefficient	Precision	Recall	
C1	0.806	0.717	0.785	0.705	0.743
C2	0.843	0.772	0.721	0.805	0.761

Despite the improvement, the model can still not predict settlement points accurately. Based on the confusion matrix of the C2 result, classification results of water bodies and others are significantly more accurate than built-up and settlement classes with precision and recall values around 0.90. The low precision and recall values of settlement and built-up were caused by misclassification, where some settlement points were predicted as built-up and vice versa.

After running the random forest classifier in Google Earth Engine, the question of which variable has the most significant predictive power was raised. The variables of high importance are the drivers of the result, and their values greatly influence the value of the result. This can be seen through the information on the RF variable importance, as shown in Figure 3. At the same time, the classification result has been drawn spatially in Figure 4.

Figure 3 illustrates the significance of all 19 training features for Banyumas as the first region in building the model by highlighting the importance of each feature individually. All features give various contributions to the classification result. We found dominant variables from every group of

features (spectral reflectance, spectral indices, and texture). The NIR band made the highest contribution from the spectral reflectance group in these findings. If we omit the NIR band in our experiment, OA and Kappa Coefficient values decrease from 0.843 and 0.772 to 0.817 and 0.733. These results support our findings.

The Sentinel-1 SAR features (VV) and NDVI spectral index became the two most significant contributors to the final classification. SAR images can provide information about the structure and dielectric properties of the surface materials in most cases and the corresponding NDVI as the most applied spectral index to detect vegetation features from other features [42]. Integrating multi-source training features

could guarantee classification accuracy across many impervious landscapes. Gong [43] also showed that using optical and SAR imagery in conjunction could significantly improve land cover classification and estimate impervious surfaces. In line with the results of Osgouei's research [32], which concludes that from the characteristic TOA reflectance values of each Sentinel-2A image channel, there is an increase in reflectance in the NIR channel (B8, B8a), so it can be used to detect vegetated areas. In this research, NIR separates class 1 (others or unclassified category, mostly of vegetation) from other classes (built up, settlement, and water body). Also, the NIR band considers several spectral indices that made a higher contribution.

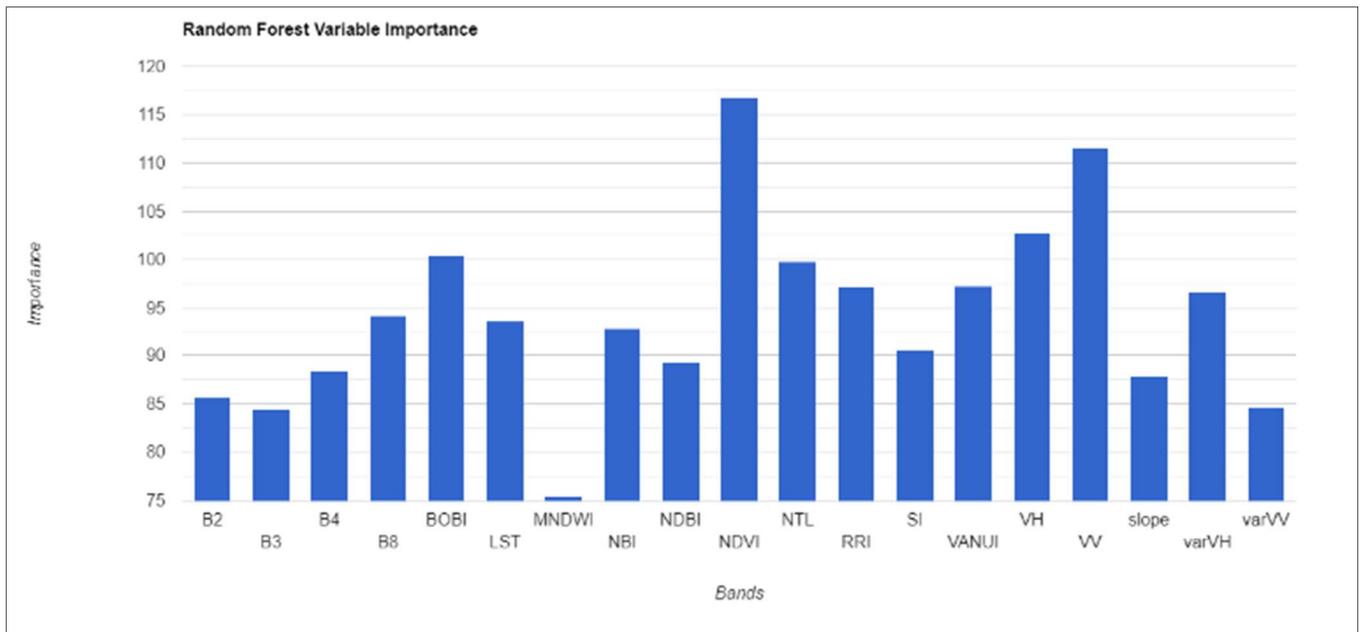


Fig. 3 The importance of the input features derived from the random forest model using the Banyumas classification model.

The classification model of four land cover classes in Banyumas was applied to other areas. Two neighboring areas, namely Tegal and Cilacap, became the destination for the pilot transfer model. By including 19 predictors, the classification model in Banyumas was able to produce slightly over the expected number, 0.843 OA (Table 3, line C2), although the separation between settlements and other built-up lands was still below 0.800. Henceforth, the transfer model results for Cilacap and Tegal regions are shown in Table 4.

TABLE IV
ACCURACY OF CLASSIFICATION MODEL AT CILACAP AND TEGAL REGENCY USING BANYUMAS TRAIN MODEL

Classification number	Validation				
	OA	Kappa coefficient	Precision	Recall	F1 score
Cilacap	0.789	0.710	0.796	0.592	0.679
Tegal	0.800	0.725	0.852	0.590	0.697

Table 4 presents the statistics for the classification model in two areas by applying the Banyumas train model. We used 19 predictor combinations. The overall accuracy of the Cilacap classification model is 0.789, with a kappa coefficient of 0.710. Meanwhile, Tegal classification results in a slightly higher overall accuracy of 0.800 with a kappa coefficient of

0.725. In line with the trained model used, the statistical results from the model transfer process have the same results as the Banyumas train model. The classification of classes outside of built-up land (water body and other) gives precision and recall values in the range of 0.788 – 1000. The case of classification between settlements and other built-up lands has not obtained the desired accuracy result (<0.800).

Through the values of precision and recall for the settlement class as stated in Table 3. It is known that the Banyumas train model can predict well the distribution of settlements in Cilacap and Tegal (precision 0.710), but the model results are sensitive in displaying according to reality. The recall value (value 0.590) indicates that the prediction results from the model are not all by the reality in the field.

The F1 score test depicted how well our model was performing. The F1 score showed the harmonic mean of the precision and recall number of the model. For both Cilacap and Tegal classification models, the F1 score falls between 0.679 – 0.697. The F-score is extensively used for examining information retrieval systems like browsers and many types of machine learning models, notably in natural language processing.

This study shows the method's ability to compile the classification model to update land cover conditions on a

massive scale. Figure 4 shows the comparison between natural features, classification results, and data on the distribution of four land cover classes from the National Base Map scale 1:25,000 (<https://tanahair.indonesia.go.id/portal-web>). The results of the model classification show that the distribution of settlements and built-up land is in line with the natural image (RGB) and is more updated than using the

information on the National Base Map available. Urban areas are detected with a lighter appearance, and the main composition is built-up land and settlements. Note that settlement class in the National Base Map is defined as settlement and activity center, while built-up class is defined as building.

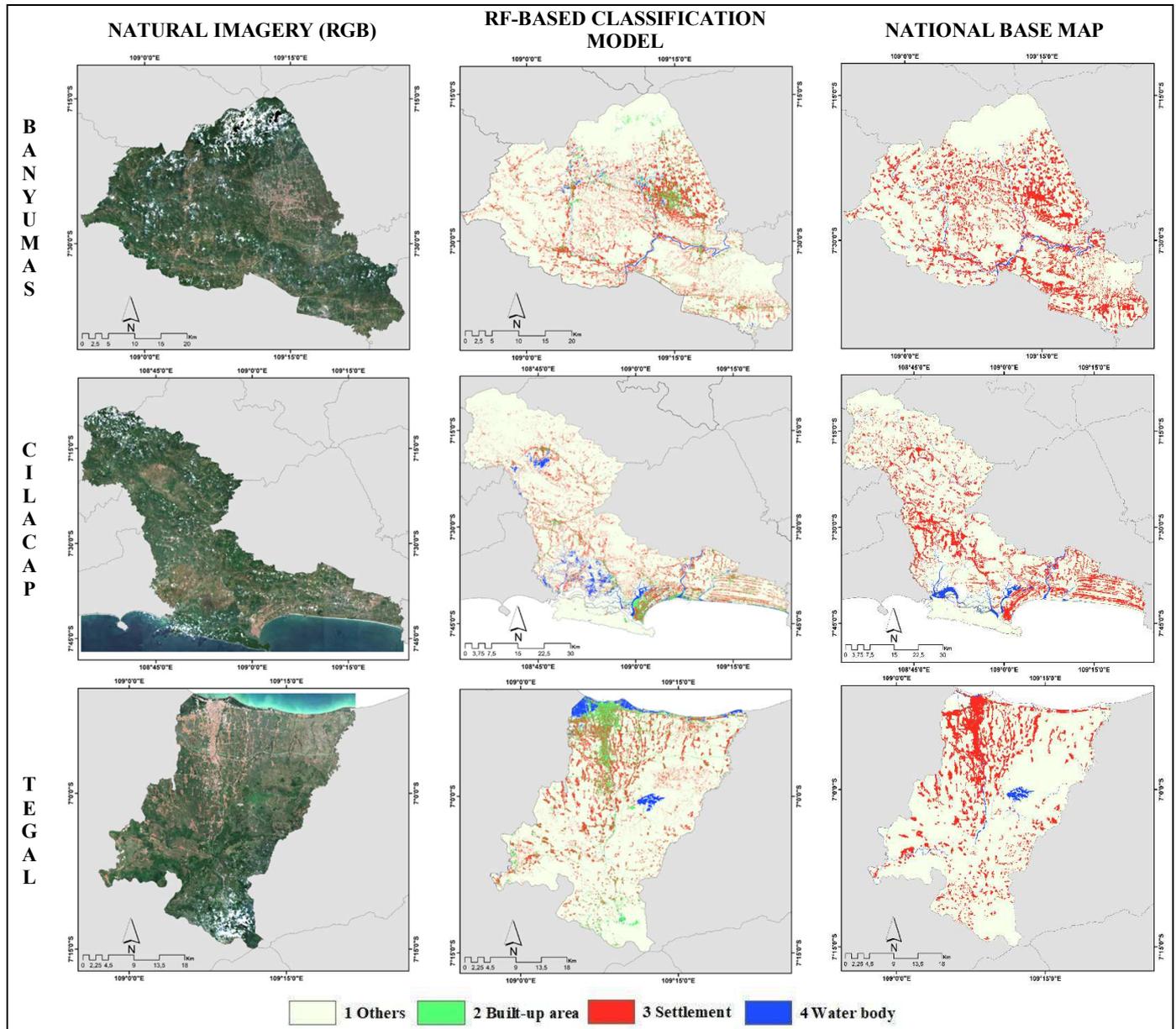


Fig. 4 Comparison between RGB image from Sentinel-2 Level 2A, classification results using RF, and land use class from the National Base Map in three study areas. Classification results of Cilacap and Tegal used the RF trained model of Banyumas.

Machine learning in separating non-residential buildings from settlements because of the similarity of the type of roof used, according to the recall and precision values obtained in the classification model in the three study areas. In line with statistical results, the trained model can generally separate other classes (which generally contain bare soil/bare land, vegetation, forestry), water classes (lakes, reservoirs, aquaculture ponds), and well-developed land. The separation between buildings and settlements is more apparent and more up-to-date (this can be seen from the RGB visual appearance,

where non-residential buildings appear lighter and whiter in color, and roads have been classified as built-up land. However, there is still confusion in some locations that are accepted).

To see the separation of the two classes of built-up land, namely non-residential buildings and residential areas, we enlarged the area on the Banyumas train model and compared it with the natural image (Figure 5). Four different points were taken to show the classifier's sensitivity in separating land cover classes. The black arrows indicate well-defined spots

between non-residential and residential buildings. Buildings and roads with a size equal to or more than the pixel size of the Sentinel-2 image can be classified well. In the natural

image, buildings are generally characterized by light colors (bright white), while brick colors characterize settlements.

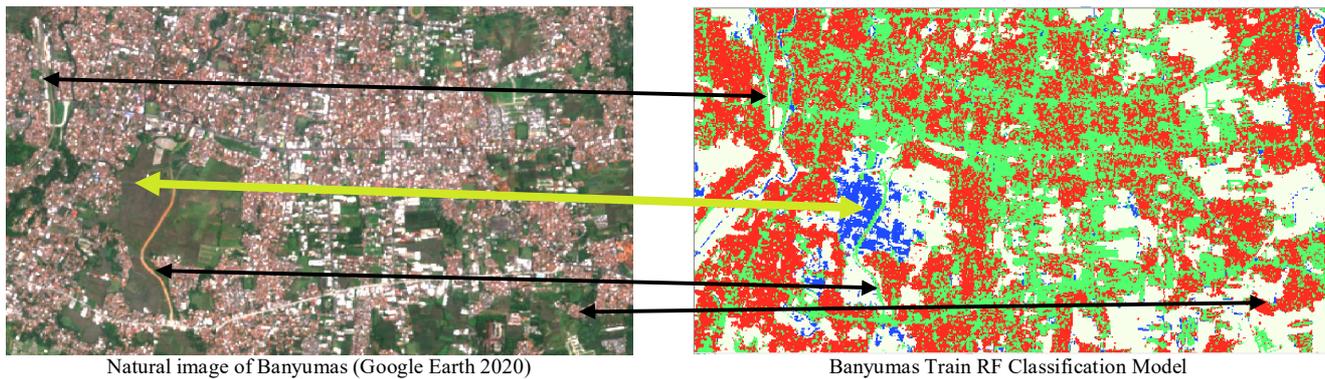


Fig. 5 Comparison between Sentinel-2 surface reflectance image using natural color composite and the classification RF result. Black arrows indicate well-defined spots between non-residential and residential buildings. The yellow arrow pinpoints the misclassification result.

In line with the results of Zheng's research [36], which combines 39 parameters from spectral features, texture features, and indices, it can identify built-up land with an accuracy above 90%. However, according to the findings in this paper, the classification is confused in buildings that have the same type of impervious surface as houses and vice versa. Furthermore, the yellow arrow indicates a classification error still experienced in the open area that should be classified as class 1 (others) instead, classified into class 4 (water body). The misclassification occurs because the rice field is still amid irrigation inundation when the image is captured.

IV. CONCLUSION

Integrating 19 predictors with the RF classifier can produce a classification model of 4 land cover classes with an overall accuracy of 0.800 in the Banyumas case. The algorithm has separated waterbody features and other classes (vegetation and open land) with up to 0.900 accuracies. Machines still experience confusion in separating settlements and other buildings due to the similarity of the impervious surface types. Transferring the Banyumas model to two neighboring areas, in a statistical test similar to the trained model, OA's accuracy value was slightly lower, 0.789 – 0.800. However, the classification method developed in this paper can assist in the rapid description of land cover when the official data sources have not provided up-to-date information yet. Too, for modest landcover mapping at the massive area in separated islands. Still, there are future challenges to developing methods that consider culture in detecting human settlements.

ACKNOWLEDGMENT

We thank the Geospatial Information Agency of Indonesia for funding this research. The paper materials were the development of SDGs research at BIG in 2021. All authors have equal contributions in this article.

REFERENCES

[1] Ministry of National Development Planning/ National Development Planning Agency, *Metadata Indikator Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs) Indonesia - Pilar Pembangunan Lingkungan*, 2nd ed. 2020.

[2] G. Rousset, M. Despinoy, K. Schindler, and M. Mangeas, "Assessment of deep learning techniques for land use land cover classification in southern new Caledonia," *Remote Sens.*, 2021, doi: 10.3390/rs13122257.

[3] Kiswanto, S. Tsuyuki, Mardiany, and Sumaryono, "Completing yearly land cover maps for accurately describing annual changes of tropical landscapes," *Glob. Ecol. Conserv.*, vol. 13, p. e00384, 2018, doi: 10.1016/j.gecco.2018.e00384.

[4] H. Schubert, M. Rauchecker, A. C. Calvo, and B. Schütt, "Land use changes and their perception in the Hinterland of Barranquilla, Colombian Caribbean," *Sustain.*, vol. 11, no. 23, pp. 1–21, 2019, doi: 10.3390/su11236729.

[5] P. S. Roy *et al.*, "Development of decadal (1985-1995-2005) land use and land cover database for India," *Remote Sens.*, vol. 7, no. 3, pp. 2401–2430, 2015, doi: 10.3390/rs70302401.

[6] E. B. Silva *et al.*, "A Definition of Visual Interpretation Criteria to Mapping Land-Use and Land-Cover in the Brazilian Biomes," *2020 IEEE Lat. Am. GRSS ISPRS Remote Sens. Conf. LAGIRS 2020 - Proc.*, vol. XLII, no. March, pp. 173–176, 2020, doi: 10.1109/LAGIRS48042.2020.9165577.

[7] G. Zhao and M. Yang, "Urban Population Distribution Mapping with Multisource Geospatial Data Based on Zonal Strategy," *ISPRS Int. J. Geo-Information*, vol. 9, no. 11, p. 654, 2020, doi: 10.3390/ijgi9110654.

[8] D. H. Bui and L. Mucsi, "From land cover map to land use map: A combined pixel-based and object-based approach using multi-temporal landsat data, a random forest classifier, and decision rules," *Remote Sens.*, vol. 13, no. 9, 2021, doi: 10.3390/rs13091700.

[9] J. T. Nugroho, Zylshal, N. M. Sari, and D. Kushardono, "A Comparison of Object-based and Pixel-based Approaches for Land Use/Land Cover Classification Using Lapan-A2 Microsatellite Data," *Int. J. Remote Sens. Earth Sci.*, 2017, doi: 10.30536/ijjires.2017.v14.a2680.

[10] A. Sekertekin, A. M. Marangoz, and H. Akcin, "Pixel-based classification analysis of land use land cover using Sentinel-2 and Landsat-8 data," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2017, doi: 10.5194/isprs-archives-XLII-4-W6-91-2017.

[11] J. Ai, C. Zhang, L. Chen, and D. Li, "Mapping annual land use and land cover changes in the Yangtze Estuary Region using an object-based classification framework and landsat time series data," *Sustain.*, vol. 12, no. 2, 2020, doi: 10.3390/su12020659.

[12] N. Bashit, N. Sari Ristianti, Y. Eko Windarto, and D. Ulfiana, "The Mapping of Land Use Using Object-Based Image Analysis (OBIA) in Klaten Regency," *E3S Web Conf.*, vol. 202, pp. 1–9, 2020, doi: 10.1051/e3sconf/202020206036.

[13] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," *Remote Sens. Environ.*, vol. 205, no. November 2017, pp. 338–351, 2018, doi: 10.1016/j.rse.2017.11.024.

[14] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover

- classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2019, doi: 10.1109/JSTARS.2019.2918242.
- [15] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sensing*, 2020, doi: 10.3390/RS12152495.
- [16] X. Liao, X. Huang, and W. Huang, "ML-LUM: A system for land use mapping by machine learning algorithms," *J. Comput. Lang.*, vol. 54, no. February, 2019, doi: 10.1016/j.cola.2019.100908.
- [17] W. Mao, D. Lu, L. Hou, X. Liu, and W. Yue, "Comparison of machine-learning methods for urban land-use mapping in hangzhou city, china," *Remote Sens.*, 2020, doi: 10.3390/rs12172817.
- [18] H. chien Shih, D. A. Stow, and Y. H. Tsai, "Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping," *Int. J. Remote Sens.*, 2019, doi: 10.1080/01431161.2018.1524179.
- [19] Y. H. Tsai, D. Stow, H. L. Chen, R. Lewison, L. An, and L. Shi, "Mapping vegetation and land use types in Fanjingshan National Nature Reserve using google earth engine," *Remote Sens.*, 2018, doi: 10.3390/rs10060927.
- [20] I. Prasetyo, W. S. Pranowo, C. L. Tobing, A. Kurniawan, and T. Puliwarna, "Analisis Mangrove dari Citra Satelit sebagai Pertahanan Pantai dengan Menggunakan Pendekatan Cloud Computing," *J. Chart Datum*, vol. 7, no. 1, pp. 47–62, 2021, doi: <https://doi.org/10.37875/chartdatum.v7i1.189>.
- [21] H. Shafizadeh-Moghadam, M. Khazaei, S. K. Alavipanah, and Q. Weng, "Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors," *GIScience Remote Sens.*, vol. 58, no. 6, pp. 914–928, 2021, doi: 10.1080/15481603.2021.1947623.
- [22] J. Camilo Fagua and R. Douglas Ramsey, "Geospatial modeling of land cover change in the Chocó-Darien global ecoregion of South America; One of most biodiverse and rainy areas in the world," *PLoS One*, 2019, doi: 10.1371/journal.pone.0211324.
- [23] P. Han, Q. Zhang, Y. Zhao, and F. Y. Li, "High-resolution remote sensing data can predict household poverty in pastoral areas, Inner Mongolia, China," *Geogr. Sustain.*, vol. 2, no. 4, pp. 254–263, 2021, doi: 10.1016/j.geosus.2021.10.002.
- [24] B. Ghansah, C. Nyamekye, S. Owusu, and E. Agyapong, "Mapping flood-prone and Hazards Areas in rural landscape using Landsat images and random forest classification: A case study of Nasia watershed in Ghana," *Cogent Eng.*, vol. 8, no. 1, 2021, doi: 10.1080/23311916.2021.1923384.
- [25] S. e. hyde. Soomro *et al.*, "Mapping flood extend and its impact on land use/land cover and settlements variations: a case study of Layyah District, Punjab, Pakistan," *Acta Geophys.*, 2021, doi: 10.1007/s11600-021-00677-4.
- [26] T. N. Widodo, H. Zubair, and R. Padjung, "Land use change study and the increased risk of floods disaster in Jeneberang watershed at Gowa Regency, South Sulawesi, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 824, no. 1, 2021, doi: 10.1088/1755-1315/824/1/012045.
- [27] A. Rudiastuti, N. M. Farda, and D. Ramdani, "Mapping built-up land and settlements: a comparison of machine learning algorithms in Google Earth engine," in *Proc. SPIE 12082, Seventh Geoinformation Science Symposium 2021*, 2021, vol. 1208206, no. December 2021, p. 47, doi: 10.1117/12.2619493.
- [28] H. Ji, X. Li, X. Wei, W. Liu, L. Zhang, and L. Wang, "Mapping 10-m Resolution Rural Settlements Using Multi-Source Remote Sensing Datasets with the Google Earth Engine Platform," 2020.
- [29] X. Zhang *et al.*, "Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform," *Earth Syst. Sci. Data Discuss.*, 2020, doi: 10.5194/essd-2019-200.
- [30] A. Al Kafy *et al.*, "Remote sensing approach to simulate the land use/land cover and seasonal land surface temperature change using machine learning algorithms in a fastest-growing megacity of Bangladesh," *Remote Sens. Appl. Soc. Environ.*, vol. 21, no. August 2020, p. 100463, 2021, doi: 10.1016/j.rsase.2020.100463.
- [31] Z. Sun *et al.*, "A review of Earth Artificial Intelligence," *Comput. Geosci.*, vol. 159, no. August 2021, p. 105034, 2022, doi: 10.1016/j.cageo.2022.105034.
- [32] P. E. Osgouei, S. Kaya, E. Sertel, and U. Alganci, "Separating built-up areas from bare land in mediterranean cities using Sentinel-2A imagery," *Remote Sens.*, vol. 11, no. 3, 2019, doi: 10.3390/rs11030345.
- [33] A. Yong and X. Bin, "Information Extraction of Urban Expansion Based on Remote Sensing A case of Jinghui irrigation district in Shaanxi Province, China," *Water Resour. Environ. Prot.*, no. 4, pp. 2683–2686, 2011.
- [34] T. Yugang, X. U. Yun, and Y. Xiaonan, "Perpendicular impervious index for remote sensing of multiple impervious surface extraction in cities," *Acta Geod. Cartogr. Sin.*, vol. 46, no. 4, p. 468, 2017.
- [35] M. Yuhe, Z. Mudan, Z. Peng, and W. Jian, "Comparison of Impervious Surface Extraction Index Based on Two Kinds of Satellite Sensors," *Spacecr. Recover. Remote Sens.*, vol. 42, no. 2, pp. 139–151, 2021, doi: 10.3969/j.issn.1009-8518.2021.02.016.
- [36] Y. Zheng, L. Tang, and H. Wang, "An improved approach for monitoring urban built-up areas by combining NPP-VIIRS night-time light, NDVI, NDWI, and NDBI," *J. Clean. Prod.*, vol. 328, 2021, doi: 10.1016/j.jclepro.2021.129488.
- [37] I. N. Hidayati, R. Suharyadi, and P. Danoedoro, "Developing an Extraction Method of Urban Built-Up Area Based on Remote Sensing Imagery Transformation Index," *Forum Geogr.*, vol. 32, no. 1, pp. 96–108, 2018, doi: 10.23917/forgeo.v32i1.5907.
- [38] F. Puturuho, P. Danoedoro, J. Sartohadi, and D. Srihadmoko, "The Development of Interpretation Method For Remote Sensing Imagery In Determining The Candidate of Landslide In Leitimur Paninsula, Ambon Island," *J. Ilmu Lingkungan*, vol. 15, no. 1, p. 20, 2017, doi: 10.14710/jil.15.1.20-34.
- [39] Y. Zheng, L. Tang, and H. Wang, "An improved approach for monitoring urban built-up areas by combining NPP-VIIRS night-time light, NDVI, NDWI, and NDBI," *J. Clean. Prod.*, vol. 328, p. 129488, 2021, doi: 10.1016/j.jclepro.2021.129488.
- [40] Q. Zhang, C. Schaaf, and K. C. Seto, "The Vegetation adjusted NTL Urban Index: A new approach to reduce saturation and increase variation in night-time luminosity," *Remote Sens. Environ.*, vol. 129, pp. 32–41, 2013, doi: 10.1016/j.rse.2012.10.022.
- [41] L. Pasqualini and M. Parton, "Pseudo Random Number Generation: A Reinforcement Learning approach," *Procedia Comput. Sci.*, vol. 170, no. 2019, pp. 1122–1127, 2020, doi: 10.1016/j.procs.2020.03.057.
- [42] S. Guha and H. Govil, "Land surface temperature and normalized difference vegetation index relationship: a seasonal study on a tropical city," *SN Appl. Sci.*, vol. 2, no. 10, pp. 1–14, 2020, doi: 10.1007/s42452-020-03458-8.
- [43] P. Gong, X. Li, and W. Zhang, "40-Year (1978 – 2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing," *Sci. Bull.*, vol. 64, no. 11, pp. 756–763, 2019, doi: 10.1016/j.scib.2019.04.024.