JOURNAL OF THE BULGARIAN GEOGRAPHICAL SOCIETY

Journal of the Bulgarian Geographical Society 53: 119–138 (2025)
DOI: 10.3897/jbgs.e155799

Research Article

Landslide exposure analysis by utilizing big geodata in Bogor area, West Java Province of Indonesia

Astisiasari Astisiasari[1], Wisyanto Wisyanto[1], Dian Melati[1], Sukristiyanti Sukristiyanti[1], Raditya Umbara[1], Yukni Arifianti[1], Trinugroho Trinugroho[1,2], Lian Andikasari[1], Taufik Ramdhani[3]

1 *Research Center for Geological Disaster, National Research and Innovation Agency (BRIN), South Tangerang, Indonesia*
2 *Faculty of Geography, University of Melbourne, Melbourne, Australia*
3 *Faculty of Engineering, University of Indonesia, Depok, Indonesia*
Corresponding author: Dian Melati (dian025@brin.go.id)

Abstract

Landslide exposure is an adept approach to measuring the consequences of a landslide hazard on elements at risk. Landslides in the Bogor area, Province of West Java, Indonesia, have increased in number and consequences since 2015. The Bogor area also has a fairly large population that may aggravate the impact. Accordingly, this study aims to measure the landslide exposures over two substantial elements (i.e., population and land use). The actual resources for these exposed elements are available from the open-access geodata and the Google Earth Engine (GEE) platform. For the land use classification, this study employs two robust machine-learning (ML) algorithms on a GEE-based Object-based Image Analysis (OBIA), i.e., Random Forest (RF) and Gradient Tree Boosting (GTB). Moreover, the population data were retrieved from WorldPop estimates. Landslide exposures were then analyzed through an overlay between these two elements with a landslide hazard map sourced from the former study. The results show that in 2020, the Bogor area had an exposed population of 885,353 people, with Bogor Selatan District having the highest exposed population (135,475 people). Moreover, in 2021, the Bogor area had a total exposed land use reaching 347.9 km^2, with built-up area having the most extensive, reaching 45.9% of the total exposed area. Here, Sukajaya District had the largest exposed land use (39.1 km^2). This study is expected to reach multifaceted entities that contribute to strengthening landslide risk reduction. Through this spatial awareness of the highly exposed areas to landslides, mitigation measures can be taken accordingly.

Key words: Element at risk, Google Earth Engine, land use, mass movement, population, Sentinel-2, WorldPop

OPEN ACCESS

Academic editor: Alexander Gikov
Received: 15 April 2025
Accepted: 19 September 2025
Published: 24 October 2025

Citation: Astisiasari A, Wisyanto W, Melati D, Sukristiyanti S, Umbara R, Arifianti Y, Trinugroho T, Andikasari L, Ramdhani T (2025) Landslide exposure analysis by utilizing big geodata in Bogor area, West Java Province of Indonesia. Journal of the Bulgarian Geographical Society 53: 119–138. https://doi.org/10.3897/jbgs.e155799

Copyright: © Astisiasari Astisiasari et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

1. Introduction

Landslides in Bogor area, Province of West Java, Indonesia, have increased significantly, from 20 in 2015 to 241 in 2021 (BNPB 2025). Since 2015, this Province of West Java has recorded the second-highest number of landslides in Indonesia with 2,205 events, at a toll of more than 17 trillion IDR (BNPB 2025). Bogor area that includes Bogor City and Bogor Regency, has had the highest

number of landslides, at 34% of the total landslides in the West Java Province (BNPB 2025). Landslides in Bogor area have caused losses of more than 250 fatalities, 328 injuries, nearly 24,000 evacuations, and more than 3,200 heavily destructed houses over the past ten years (BNPB 2025).

Generally, landslides may cause buildings and infrastructure damages, land use disturbance, and health issues (Berov et al. 2024). To mitigate these adverse losses, a pivotal risk assessment shall be conducted to analyze the hazard, vulnerability, exposure, and capacity aspects (UNISDR 2017). Concerning this, the exposure analysis quantifies the valuation of elements at risk (Mallick et al. 2021; Caleca et al. 2024), that have potential loss from a hazard. It contains the necessary environmental, physical, economic, and social factors for measuring the loss of elements at risk (Garcia et al. 2016; Mallick et al. 2021; Rahman et al. 2022).

Bogor area has a fairly large population, with more than 6.6 million people in 2021 (5.5 million in Bogor Regency and 1.1 million in Bogor City) (BPS 2022a, b). With this vast population and landslide events, Bogor area is prone to the risk of landslides. It leads to the need to study the landslide consequences over the important elements at risk, as it has not yet been conducted in the Bogor area. Related to this, spatial analysis and techniques applications (i.e., Geographic Information System, earth observation, and remote sensing) have been widely utilized in disaster management, to effectively visualize and deliver important aspects in disaster risks, including hazard, vulnerability, and exposure (Sarafova 2022, 2023).

In this present study, to measure the exposed population in a landslide hazard zone, the population dataset shall be distributed as a settlement resident, disaggregated from an administrative unit. This distributed population can be retrieved from the WorldPop data, in which annual population numbers are gridded at 3-arc (around 100-m spatial resolution), by using a Random Forest approach (Gaughan et al. 2013; Lloyd et al. 2019). Therein, the gridded data for Indonesia was initially sourced from the 2020 population census and adjusted to match the UNDP estimate (WorldPop 2025).

Moreover, to measure the actual exposed land use, the available geospatial information has various resources from the advancing geodata (Goodchild 2018). Big geodata has various spatial-temporal resolutions in near real-time (NRT) products at no cost. Here, the Google Earth Engine (GEE) accommodates a cloud computing platform to rapidly access and analyze big geodata (Gorelick et al. 2017).

Through GEE, land use classification can be generated from the available satellite image through two frequent approaches, i.e., pixel-based and object-based image classifications. In a high spatial resolution image, an object may be present in several pixels, producing a significant spectral variability due to the complexity of spectral responses (Blaschke et al. 2014). In this case, pixel-based classification may find a misclassification when a pixel informs dissimilar objects from its neighboring pixels, even though they are the same objects. Therefore, object-based image analysis (OBIA) for high spatial resolution performs better accuracy compared to the pixel-based approach (Blaschke 2010). Accordingly, this study applies a GEE-based OBIA through two machine-learning (ML) algorithms for image classification, so-called ensemble decision tree-based classifiers, i.e., Random Forest (RF) and Gradient Tree

Boosting (GTB). The ensemble of decision trees has shown better accuracy than a single decision tree (Müller and Guido 2016).

Accordingly, this study aims to analyze the landslide exposures on two substantial elements (i.e., population and land use) in Bogor area by utilizing big geodata. This objective is accomplished through: 1) classifying the element at risk of land use through MLs-based OBIA (i.e., RF and GTB), by utilizing the geodata platform of GEE, 2) estimating the exposed land use area in the landslide hazard zone, and 3) estimating the exposed population number from the gridded population model of WorldPop. This landslide exposure study is an important aspect in landslide risk analysis, as it provides pivotal information on the possible exposed elements in hazard zones. Thus, contributions to mitigation measures can be taken accordingly.

This study serves as a significant precursor to landslide risk assessment in measuring the probable exposed elements in hazard zones. It demands an actual and high-quality dataset that may require an extensive budget. Correspondingly, this study highlights the opportunity to utilize big geodata in establishing the elements at risk datasets. In the Bogor area, valuable knowledge on highly exposed areas to landslides has yet to be addressed. To our knowledge, studies related to landslides in the Bogor area contributed to analysing the susceptibility (Silalahi et al. 2019; Wicaksono et al. 2020; Melati et al. 2024; Umbara et al. 2024) and vulnerability (Trisnafiah et al. 2023).

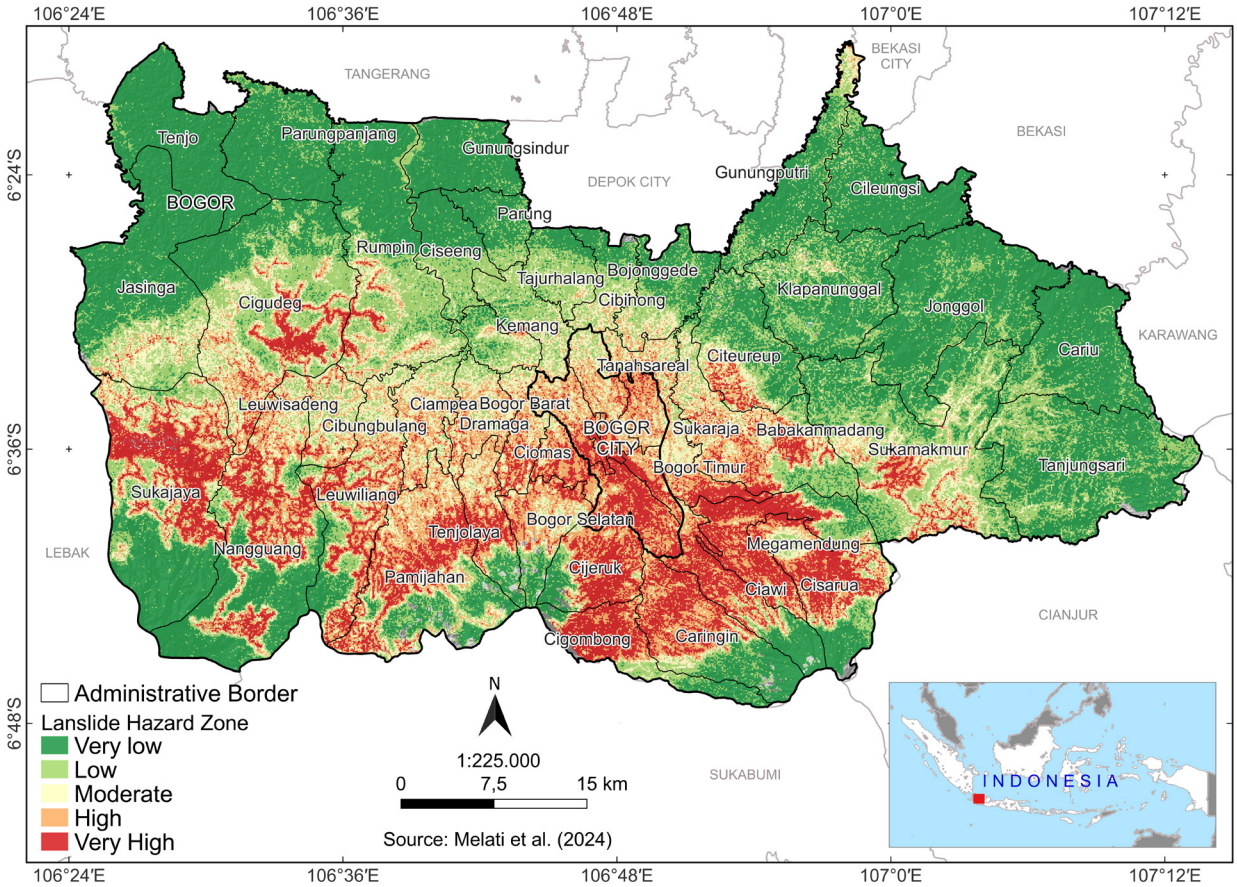

Figure 1. Study area of the landslide hazard zone in Bogor area (Melati et al. 2024).

2. Material and methods

2.1. Study area

Bogor area, including Bogor City and Bogor Regency (fig. 1), has an extent of 3,105 km^2, with a main elevation of <900 m a.s.l., and a dominant slope of 5–24° (Umbara et al. 2024). According to the Geological Map produced by the Indonesian Geological Agency, Bogor has a dominant volcanic product lithology of quaternary extrusive rock, older volcanic deposits, and intrusive rocks, with 28% sedimentary deposits and 23% alluvial deposits. In addition, Bogor area has a wet to very wet tropical climate, where almost 98% of the area has an annual rainfall of 3,000–5,000 mm/year (Umbara et al. 2024).

In 2021, Bogor City had a population density of 8,881 people/km^2. 44% of the total population was within working age, with employees as the main occupation. Bogor City had a Gross Regional Domestic Product (GRDP) of nearly 53 trillion IDR (BPS 2022b). Moreover, Bogor Regency had a population density of 1,838/km^2. 44.5% of the total population was within working age, with employees as the main occupation. Bogor Regency had a GRDP of over 267 trillion IDR (BPS 2022a).

2.2. Landslide hazard

Landslide hazard is one of the important factors in conducting landslide exposure analysis. Evaluation of hazards can be defined based on their probability, such as spatial probability, temporal probability, volume probability, or runout probability (Corominas et al. 2014). Evaluation with temporal probability often experiences constraints due to data limitations. Therefore, spatial probability in the form of landslide susceptibility was used as a final product of hazard probability (Corominas et al. 2014; Caleca et al. 2022).

This study employs former research on Landslide Susceptibility Mapping (LSM) in Bogor area, at a pixel size of 25 x 25 meters (Melati et al. 2024). The study compared seven ML-based LSM, incorporating 13 landslide causal factors and 822 landslide events between 2018 and 2022. The most reliable LSM was demonstrated by the Random Forest (RF) algorithm, achieving the highest accuracy with an Area Under Curve (AUC) value of 94.4% on the Receiver Operating Characteristic (ROC) Curve. Moreover, the accuracy and precision of this RF-based LSM were the highest among other models, reaching 0.868 and 0.861, respectively.

In addition, the LSM map, including the one utilized in this study, accommodates some degree of uncertainty (van Westen et al. 2008; Reichenbach et al. 2018). It may stem from various factors, e.g., the landslide dataset is featured in points rather than polygons, variation in the spatial resolution of landslide causative factors, and the acquisition of the non-landslide dataset lacks any standardized procedure.

Fig. 1 shows the Landslide Susceptibility Zone (LSZ) in the study area, which is divided into five classes, i.e., very low (38.3%), low (20.9%), medium (14.6%), high (13.5%), and very high (12.7%). In this study, the landslide-prone areas were explicitly opted for the high and very high susceptibility classes as the danger pixels. A concept of danger pixel, together with land use data, is utilized as a semi-quantitative method to perform a landslide risk assessment (Mallick et al. 2021; Riaz et al. 2023).

2.3. Elements at risk dataset

2.3.1. Population dataset

The population data is sourced from the WorldPop which provides a wide variety of gridded population numbers by using an RF-based approach (Stevens et al. 2015). WorldPop is freely accessible from the repository hub (WorldPop 2025), at a spatial resolution of 3 and 30 arcs (around 100 m and 1 km at the equator), in a geographic coordinate system (WGS84). WorldPop has three types of disaggregation methods, including unconstrained, constrained, and United Nations (UN)-adjusted population estimates (WorldPop 2025).

This study utilized the 2020 WorldPop dataset of the constrained with UN-adjusted population at a 100-m resolution. This data type estimates the settlement inhabitants within a 100-m pixel resolution and is adjusted to the population estimation from the Department of Economic and Social Affairs of the UN Secretariat (Bondarenko et al. 2020). Here, the gridded data for Indonesia was initially sourced from the 2020 population census and adjusted to match the UNDP estimate (WorldPop 2025).

The WorldPop dataset relies on datasets built from census statistics and satellite imagery. However, the WorldPop dataset also highlights certain limitations, such as the unavailability of real-time population counts and the insufficiency of the population training dataset. Moreover, the accuracy of building footprints is sensitive to the population estimation. Whereby, incorrect buildings may lead to population undercount, and overlooked building footprints may cause overcount population (WorldPop 2025).

2.3.2. Land use dataset

To derive the landuse dataset, this study utilized a Level 2A Sentinel-2 image that was acquired and processed in GEE. Sentinel-2 data consists of 13 spectral bands, with spatial resolutions varying at 10 m, 20 m, and 60 m. Four bands are available at a 10 m resolution, including B2, B3, B4, and B8. Six bands at 20 m are B5, B6, B7, B8A, B11, and B12. And three bands at 60 m are B1, B9, and B10 (ESA 2015). The images were filtered by dates to obtain the least cloud coverage, resulting in the 2021 acquisition. A median value was then calculated for each pixel of the collected images. And to apply OBIA, two steps were performed, including image segmentation and image classification (Lillesand et al. 2015), which are further explained below.

2.3.2.1. Image segmentation

Image segmentation aims at grouping neighboring pixels with similar characteristics into a meaningful object (Marpu et al. 2010). An algorithm of Simple Non-Iterative Clustering (SNIC) provided in GEE was used to conduct image segmentation. Our study performed SNIC using Sentinel-2 images for the bands of blue, green, and red. Some parameter settings need to be adjusted to implement the SNIC algorithm (Achanta and Susstrunk 2017), including compactness, connectivity, neighborhoodSize, and seeds. Compactness determines the cluster form, connectivity performs the proximity to combine clus-

ters, neighborhoodSize prevents the artifacts on tile boundary, while seeds determine the size of the cluster.

2.3.2.2. Image classification

For image classification, a sample dataset was collected by establishing a systematic point at a 5 km distance. At each point covering a 200 x 200 m area, a visual interpretation was performed. Moreover, a sample dataset was also generated outside the systematic points for additional land use data. In total, the sample dataset consists of approximately 2000 polygons. This sample dataset was then divided into 70% training data and 30% validation data.

Sentinel-2 with a true-color composite visualization and Google Earth image with higher spatial resolution were utilized to visually guide the land use interpretation. Six land uses were classified, including cropland, built-up area, open area, shrubs, water body, and forest. Whereas cloud cover is also classified but excluded from further analysis. The following key elements were used to guide the visual interpretation:

- Croplands for agricultural fields are shown in a light green tone, smooth texture, and regular pattern.
- Built-up areas for settlements are represented with a range of hues from white, grey, light to dark brown, and irregular patterns. While factory buildings appear in white, coarse texture, and regular pattern.
- Open area is indicated with a varied color from white, light yellow to brown, irregular pattern, and smooth texture.
- Shrubs are shown in green tone, irregular pattern, and coarse texture.
- Water bodies are depicted in dark blue, an irregular pattern, and smooth texture.
- Forests are shown in dark green, an irregular pattern, and coarse texture.

Moreover, this study utilized 15 predictor variables to classify the image. Spectral bands of the Sentinel-2 Multispectral Imaging (MSI) were used, consisting of Band-2 (blue), Band-3 (green), Band-4 (red), Band-5 (red edge(RG)-1), Band-6 (red edge(RG)-2), Band-7 (red edge(RG)-3), Band-8 (NIR), Band-8A (red edge(RG)-4), Band-11 (SWIR-1), and Band-12 (SWIR-2). In addition, five spectral indices were also used as predictor variables, as can be seen in Eqs 1–5. The Normalized Difference Vegetation Index (NDVI) is used to enhance the vegetation cover. Both NDVI RE1 and NDVI RE2 handle the lower chlorophyll absorptions compared to the red band, thus reducing the saturation effect (Gitelson and Merzlyak 1996). The Normalized Difference Water Index (NDWI) and Visible red/NIR-based Built-up Index (VrNIR-BI) are used to enhance the water body (McFeeters 1996) and built-up area (Estoque and Murayama 2015), respectively.

$$NDVI = (NIR - R) / (NIR + R) \quad \text{(Eq. 1)}$$

$$NDVI\ RE1 = (RE2 - RE1) / (RE2 + RE1) \quad \text{(Eq. 2)}$$

$$NDVI\ RE2 = (RE3 - RE1) / (RE3 + RE1) \quad \text{(Eq. 3)}$$

$$NDWI = (GREEN - NIR) / (GREEN + NIR) \quad \text{(Eq. 4)}$$

VrNIR-BI = (RED - NIR) / (RED + NIR) (Eq. 5)

With all of the predictor variables, image classification was further conducted using RF and GTB algorithms provided by GEE. These two types of ensemble decision trees are machine-learning algorithms that integrate many decision trees. The method can act as a classifier and regressor. The ensemble method is assumed to be more robust than a single decision tree, indicated by its better predictive performance (Sahin 2020). The ensemble method is preferred for managing complex datasets. However, there are some deficiencies regarding computational efficiency and interpretability. The single decision tree is much lighter than the ensemble, as the computational process is much faster. Using shallow decision trees, we can interpret the classifier easily. In contrast, the ensemble method is frequently less transparent, making it difficult to understand. Nonetheless, the visualization of the tree may help to understand the model (Sahin 2020). Even though both of them are ensemble techniques, they have their own characteristics.

RF was initially introduced by Breiman (2001). RF obtained its name by involving randomness in the building of each tree to ensure that every tree is not similar. The result of the model prediction is assigned through the majority vote across all the trees. There are two ways to randomize the trees in RF: choosing the samples to build a tree and choosing the features used for the split test (Müller and Guido 2016). RF has an advantage to overcome overfitting. Though each tree in RF may predict well, it may remain an overfit on a particular part. When many trees are built, all of them work well, but overfitting may occur in different ways. The overfitting can be reduced by averaging all their results.

GTB utilizes the gradient descent technique (Friedman 2002). The trees are built gradually, where each tree will correct the mistakes from the former one (Müller and Guido 2016). Compared to RF, GTB does not employ randomization in building trees; instead, it employs strong pre-pruning. Pre-prunning is conducted by setting a particular threshold during tree building to prevent building additional nodes. It is useful to reduce overfitting and complexity, as well as to improve computational efficiency. GTB often uses very shallow trees (max_depth <= 5), making the model simpler with low memory and fast prediction. GTB integrates several simple models (i.e., shallow trees). Each tree is only able to produce accurate predictions in some parts of the data; therefore, more trees are built sequentially to deliver a better prediction.

In the present study, image classification used 30% of the sample dataset as the validation data to generate the model performance. Evaluation on the resulting image classification was performed through validation metrics such as Overall Accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA), and Kappa index (Foody 2002). OA measures the percentage of objects that have been correctly classified. It is derived from the error matrix, which further calculates the UA and PA (Foody 2002). The error matrix provides a comparison between the classification result and the reference dataset used for validation. UA measures the accuracy of the classification result, while the PA evaluates the accuracy of reference objects that are classified.

Altogether, OA, UA, and PA metrics are generally reliable for accuracy assessment, as they explicitly indicate the probabilities that distinguish the data

quality of a map (Stehman 1997). Nonetheless, studies also recommend the Kappa index as the standard measure for accuracy assessment (Dash et al. 2023). It evaluates the degree of agreement between the classified and reference data by considering the potential of random agreement. It also addresses the issue of assigning the accurate class (Foody 2002). Moreover, the Kappa index also effectively demonstrates model performance with an imbalanced dataset (De la Cruz Huayanay et al. 2024; Khan et al. 2024).

2.4. Exposure analysis

Analysis of landslide exposure plays a key role in understanding and further mitigating the possible consequences of landslides on the infrastructure, population, and environment to achieve sustainable development (Mateos et al. 2020). In addition, cropland and built-up areas, as well as settlements, are essential to societies. Cropland is important to provide food for societies as well as livelihood in the agricultural sector. Landslides may also threaten biodiversity found in forests due to natural habitat disturbance (Rahman et al. 2022).

Our study quantified the degree of exposure by superimposing the extent of the elements at risk (i.e., built-up area, cropland, forest, population) and the LSZ on the Geographic Information System (GIS) environment. The exposure analysis was conducted within a pixel size of 100 m. Fig. 2 shows the flowchart of this study. Each of the exposure elements is masked for the LSZ (on high and very high classes) from the former study (Melati et al. 2024). Subsequently, the exposures to landslides are performed at each district-level unit.

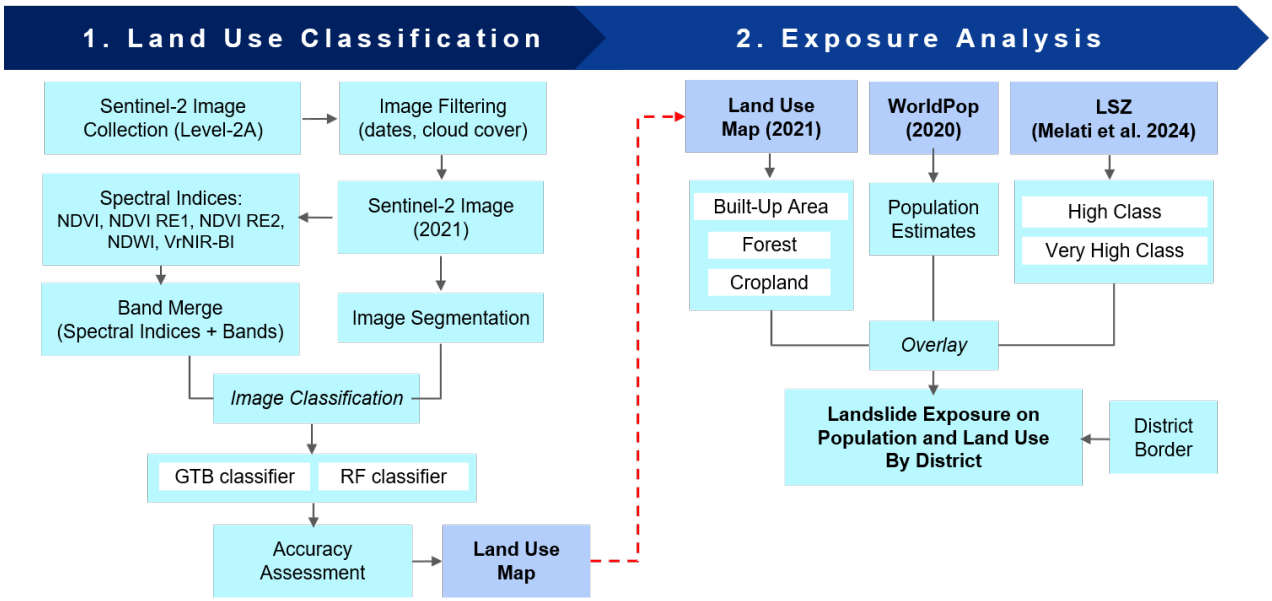

Figure 2. Flowchart of the study on exposure analysis.

3. Results

3.1. Land use mapping

Image segmentation was performed by implementing the parameter settings of compactness factor = 0, connectivity = 8, neighborhoodSize = 128, and seed spacing = 10. This optimization of parameter setting has been done by visual observation following the former study in a similar landscape within Indonesia (Melati et al. 2022, 2024). Generally, different seeds generate varying amounts and object sizes as a result of image segmentation. It was tested to produce image segmentation with different seed spacing (s) (i.e., 5, 10, 15, 20, and 25). The higher the s, the larger the object size. In this case, the segmentation produces larger objects or under-segmented fragments that may lead to less accurate results for further image classification. On the contrary, the low seed spacing produced an over-segmented image, which may require more time to classify the image.

After image segmentation, image classification was then performed by utilizing RF and GTB by using the training dataset. Both models were repeatedly tested using different numbers of trees, and then chosen with the highest OA and Kappa. By using a validation dataset, the resulting RF with number of trees = 100 and GTB with number of trees = 20 were chosen and further evaluated. It was revealed that the RF algorithm produced better OA than the GTB algorithm (i.e. 0.990 vs 0.988) as well as better Kappa index (i.e. 0.987 vs 0.984).

According to the results of UA and PA, as shown in Fig. 3, it is seen that both classifiers perform excellently with a value > 90%. However, RF outperforms GTB classifier, where the values of UA and PA of each class are higher for the RF classifier. Both classifiers tend to have similar abilities in classifying the objects. For instance, the lowest and highest values of UA and PA for both classifiers are found in the same class (i.e., shrubs and cloud cover, respectively). As RF performs higher OA, UA, and PA, we used the land use map produced by RF classifier to further identify the elements exposed by landslides.

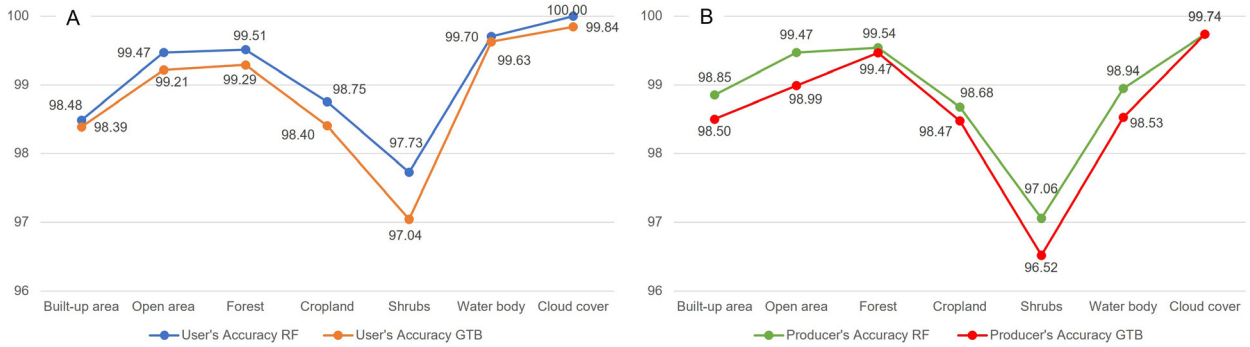

Figure 3. Accuracy produced by RF and GTB classifiers. **A** user's accuracy **B** producer's accuracy.

In our study, RF classifier performs well for the classes of the selected elements at risk. The UA values of built-up area, forest, and cropland are 98.48%, 99.51%, and 98.75%, respectively. Furthermore, the PA values are 98.85%, 99.54%, and 98.68%, respectively. The confusion between classes can be seen in Table 1. Regarding the classification result, the built-up area is mostly confused with the cropland. Forest and cropland classes are mostly confused

with built-up areas. In addition, shrubs are mostly confused with built-up area. Even though some confusion among classes occurred, it can be seen that the classifiers performed well, as the PA and UA of each class are high.

Table 1. Error matrix of image classification by RF classifier.

		Classification						
		Built-up area	Open area	Forest	Cropland	Shrubs	Water body	Cloud cover
Reference	Built-up area	23,121	6	84	73	103	3	0
	Open area	17	6,591	7	7	2	2	0
	Forest	74	14	34,349	41	30	0	0
	Cropland	130	6	42	15,130	19	6	0
	Shrubs	115	1	29	60	6,797	1	0
	Water body	19	7	3	10	4	4,029	0
	Cloud cover	1	1	3	0	0	0	1,897

3.2. Exposure analysis

3.2.1. Population exposure

In 2020, Bogor area had an overall exposed population to landslides at 885,353 people across 34 districts. Fig. 4 depicts the map of exposed population by district in Bogor area in 2020. From the total of 46 districts in Bogor area, the population of 12 districts is considered unexposed to landslides. These districts are mainly located in the very low to low landslide hazard zone, spread over the northern and northeastern parts of Bogor area. Even though the two unexposed Districts of Gunungputri and Cileungsi were among the three most populated districts in Bogor area (BPS 2022a), being situated in the low and very low landslide hazard zones renders them unexposed.

Fig. 5 shows that the District of Bogor Selatan had the highest exposed population of 135,475 people, or 15% of the total exposed population. It is followed by the 12% exposure in the District of Cisarua (108,401 people), 9% in the District of Megamendung (76,159 people), and 6% in both Districts of Bogor Barat and Ciawi (55,399 and 51,368 people, respectively). For the remaining 29 districts, each has less than 6% of the total exposed population.

3.2.2. Land use exposure

Fig. 6 shows the exposed land use in Bogor area in 2021. It shows that the built-up area had the most extensive exposure for nearly 160 km^2, followed by the forest at 120.8 km^2, and cropland at 67.5 km^2. Moreover, three land uses of shrubs (22.6 km^2), open area (3 km^2), and water (1 km^2) have relatively minor coverages and are considered unexposed to landslides.

Fig. 7 depicts the map of landslide exposure on land use area (km^2) by district. From the total of 46 districts, 12 districts have unexposed land use areas for having very low to low classes of landslide hazard zones. Fig. 8 presents the exposed land use area in 34 districts, at a total of 347.9 km^2. The graph shows the District of Sukajaya as the most extensive exposed land use at 39.1 km^2,

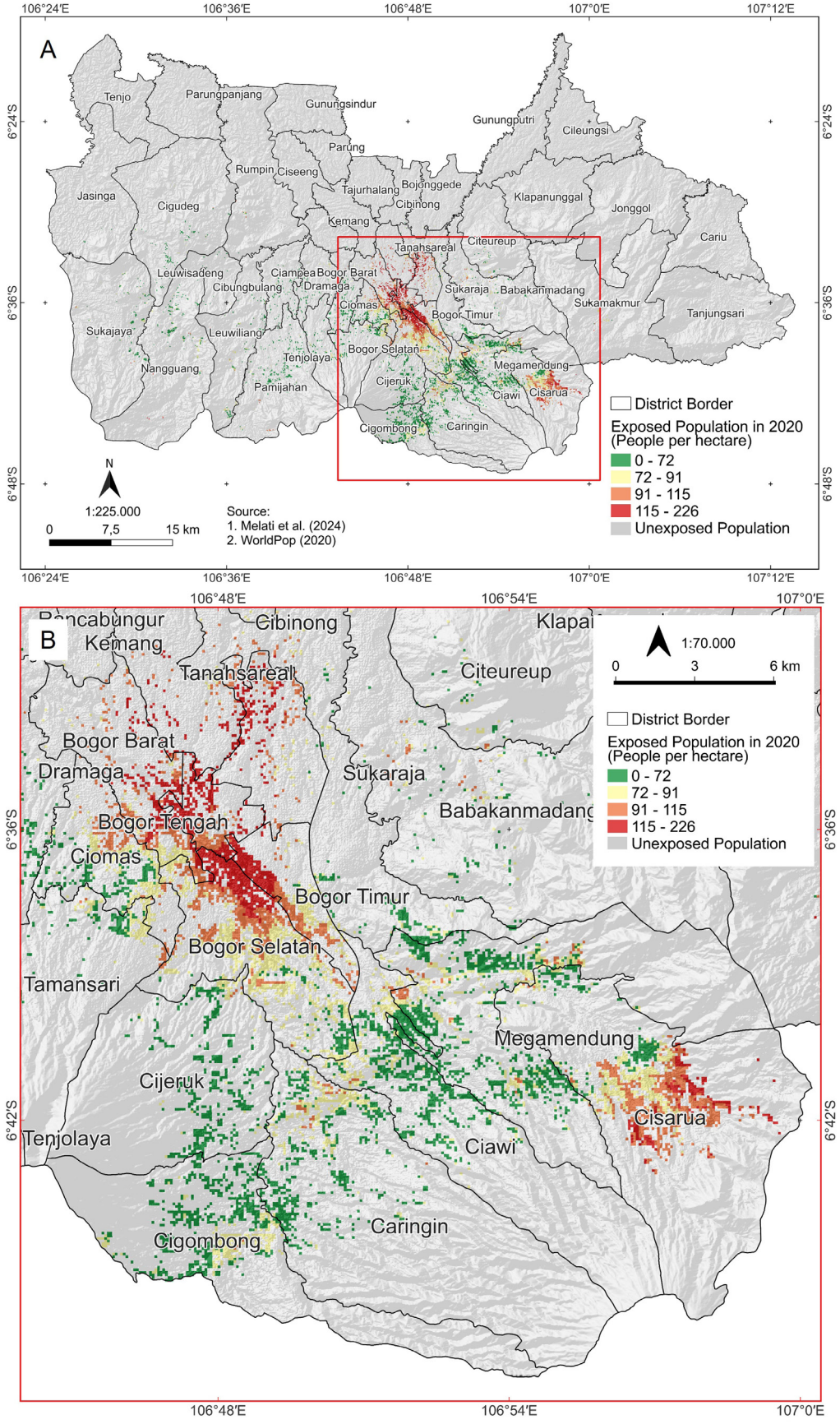

Figure 4. Map of exposed population year 2020. **A** in Bogor Area, where the highest exposed area is in the red box **B** zoomed in section of the exposed area.

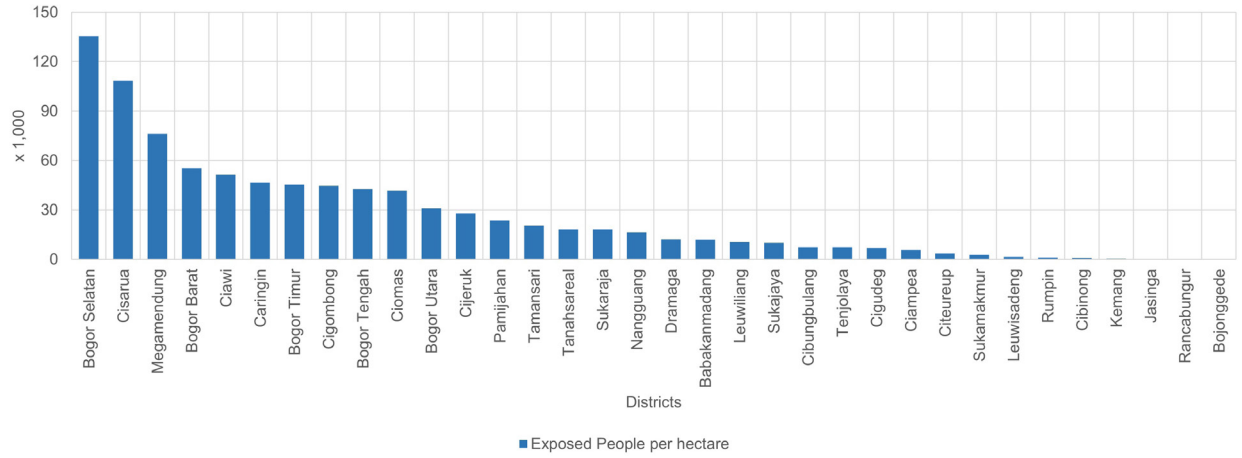

Figure 5. Histogram of landslide exposure on population by districts in Bogor Area, year 2020.

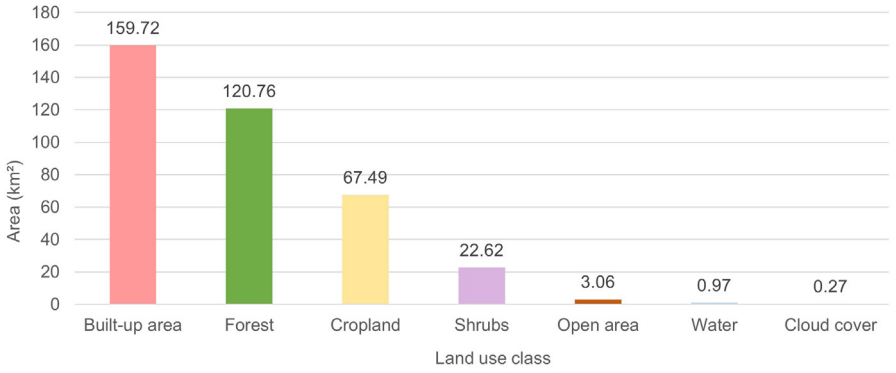

Figure 6. Landslide exposure on land use (km^2) in Bogor Area, year 2021.

followed by the Districts of Nangguang (29.5 km^2), Pamijahan (28.1 km^2), and Megamendung (25.7 km^2). For the remaining 30 districts, each has less than 25 km^2 of the total exposed land use area (or <7% of the total exposed land use).

Regarding the types of exposed land use, the three highest-exposed districts of Sukajaya, Nangguang, and Pamijahan have the highest exposure to forest, and the district of Megamendung has built-up area as the most exposed one. Moreover, of the exposed built-up area per se, the District of Bogor Selatan had the largest area at 14.2 km^2. While in the exposed forest, the District of Sukajaya had the most extensive exposure at 21.3 km^2. And for the cropland, the District of Sukajaya had the highest at 10.7 km^2.

4. Discussion

The characteristics of rapid urbanization within tropical countries (i.e., land use change, urban sprawl, and population increase) may intensify the risk of landslide hazard (Ozturk et al. 2022). Concerning the high vulnerability of economic, social, political, and cultural aspects in the Global South, landslides have had a remarkable impact, particularly in the developing tropical region (Quesada-Román and Campos-Durán 2023). Such a high intensity of structure exposure, perceived as built-up area in our study, signifies a considerable impact on key economic sectors (Berov et al. 2024).

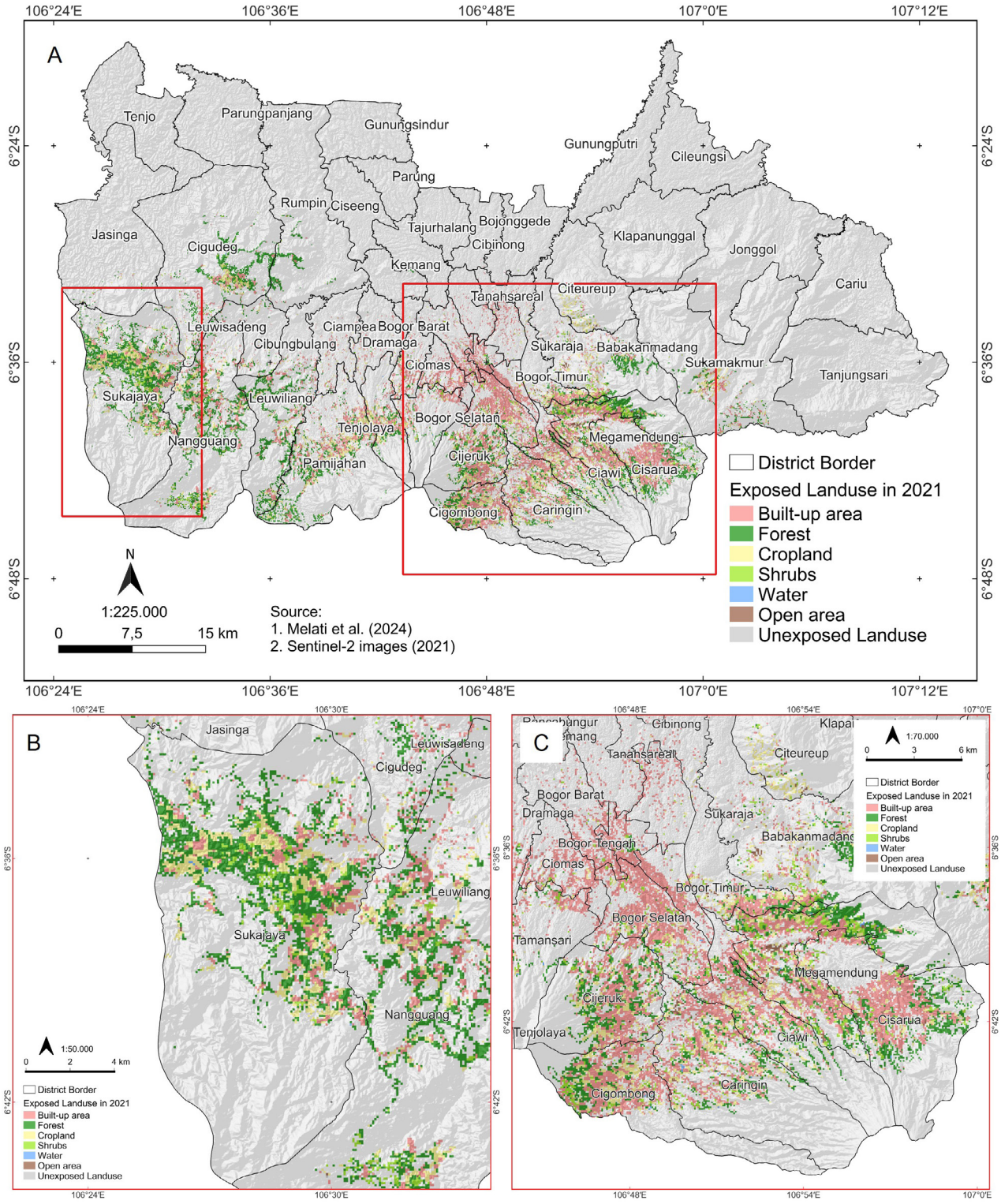

Figure 7. Map of exposed land use, year 2021. **A** in Bogor Area, where the highest exposed area is in the red box **B** zoomed in section-1 of the exposed area **C** zoomed in section-2 of the exposed area.

Nonetheless, the study on landslide exposures remains modest, as most landslide studies on the susceptibility mapping lack oversight of the exposed elements at risk (Luu et al. 2024). Analyzing landslide exposure is critical to mitigating landslide risk and protecting communities and infrastructure (Kjekstad and Highland 2009). However, risk assessment in our study area, which is

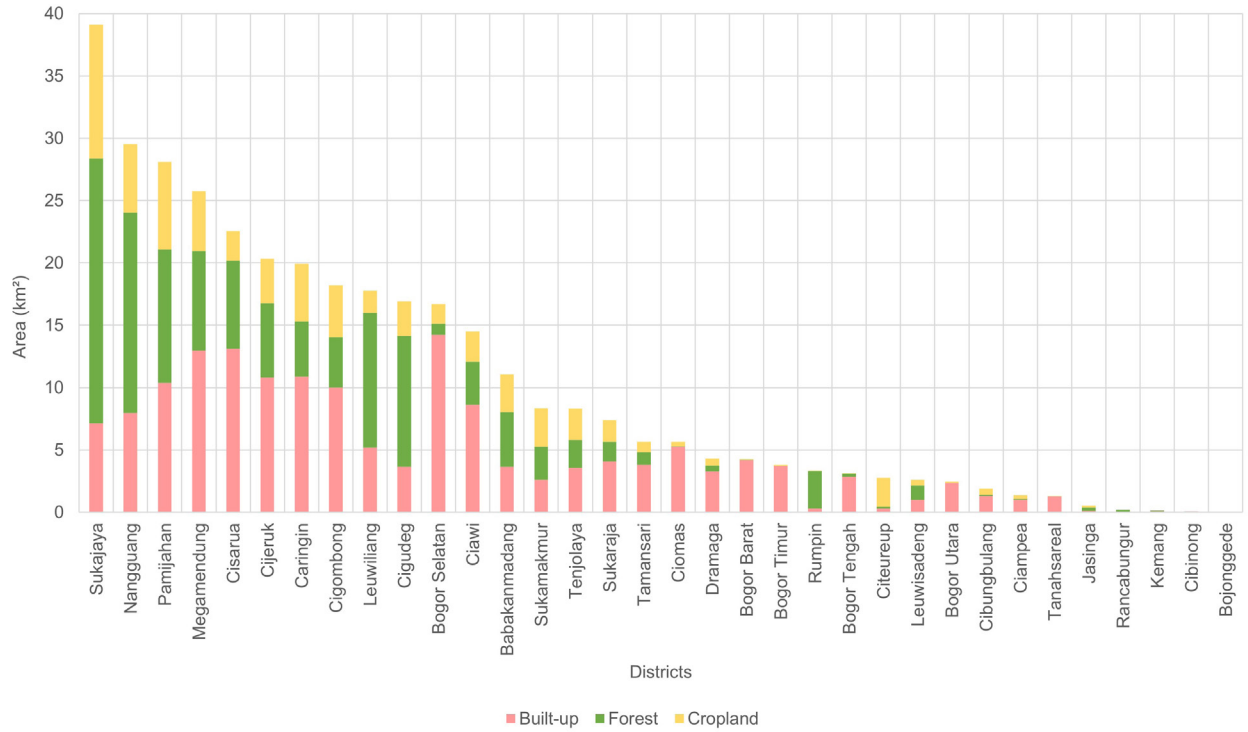

Figure 8. Histogram of landslide exposure on land use area (km²) by districts in Bogor Area, year 2021.

located in a tropical region, posed the scarcity of systematic data collection. The economic situation is the main constraint, as the majority of tropics regions are developing countries (Amarasinghe et al. 2024). Therefore, the current development of freely accessible big geodata enables users to generate respective information regularly at no cost. In case of exposure monitoring, big geodata provides further advantage to updating the exposed resources (e.g., land use and population) as the catalog is repeatedly updated (Gorelick et al. 2017).

Within the emerging era of big data, former studies on landslide exposure addressed a range of challenging gaps with adaptive approaches. The availability of open datasets can contribute to carrying out the exposure analysis (Galderisi and Limongi 2021). In a comparable context of low- to middle-income countries, the limitation of a proper database also applies in Cuenca City, Ecuador. And thus, proxy data on population are retrieved from the available open-access source (Di Napoli et al. 2023). In Kuala Lumpur, Malaysia, the disaster exposure databases are retrievable from five sources, but these datasets are not applicable for local-scale analysis (Muhamad et al. 2021).

Our study utilized big geodata to study landslide exposure within a landslide-prone area in Indonesia. The big geodata can overcome the issue of actual dataset availability by engaging the GEE platform. The GEE environment provides a large amount of satellite images, which enable the generation of existing land use datasets. Therein, further techniques on image classification are provided to deliver better accuracy on land use classification. Here, we conducted land use classification by implementing MLs-based OBIA (i.e., RF and GTB). Our results showed that RF outperforms GTB algorithm in classifying land use within our study area. This finding aligns with former studies in which RF performed with better accuracy than GTB (Ouma et al. 2023; Zhang et al. 2025).

Consequently, land use classes produced by RF image classification were used to further analyze the land use exposed by landslides within the study area.

This study has some limitations in the acquisition of available data. Firstly, the landslide susceptibility obtained from the former study utilized landslide inventory data up to 2022. Secondly, the land use dataset was generated in 2021, and lastly, the population dataset was derived in 2020. Moreover, although ground validation was not performed in land use classification, the use of Google Earth imagery can minimize the misinterpretation and has proven effective in former studies (Nasiri et al. 2022; Dash et al. 2023). Furthermore, to address spatial resolution discrepancies, the landslide susceptibility and land use datasets were resized at 100 m, following the spatial resolution of the population dataset. This analysis unit thus results in an aggregate estimation over the mixture of actual elements.

However, it should be emphasized that the findings from a relatively small-scale analysis may serve as a considerable starting point for developing nations (Caleca et al. 2024). The initial information on the spatial distribution of landslide exposures (e.g., population and land use) may further deliver insight to take mitigation measures on spatial planning. Moreover, information on potential losses is pivotal in developing resilient communities. Whereas further work in engaging the exposed elements with the involvement of multi-faceted stakeholders can safeguard local communities in the Bogor area, West Java Province of Indonesia.

5. Conclusion

This study of landslide exposure measures the possible consequences over the two substantial elements at risk, i.e., population and land use, in Bogor area, Province of West Java, Indonesia. The resulting landslide exposures were derived from the overlaid elements at risk with the landslide hazard map (on the high and very high hazard zones), sourced from the former study (Melati et al. 2024). Among the 46 districts in Bogor area, 12 districts were considered unexposed to landslides for having the very low and low classes in the hazard zones.

The results estimate that Bogor area had a total exposed population of 885,353 in 2020, and a total exposed land use of 347.9 km^2 in 2021. The District of Bogor Selatan had the highest exposed population at 135,475 people (15% of the total exposed population). While the District of Sukajaya had the largest exposed land use at 39.1 km^2 (11.2% of the total exposed land use). Moreover, built-up area was the most extensive exposure among other classes, reaching 159.7 km^2 (45.9% of the total exposed land use).

The methodological approach proposed in this study is replicable in other tropical regions and also overcomes the data-scarce areas. Regarding the land use classification, it is recommended to collect the actual field data within the available budget to ensure the independence of validation data. However, the availability of open-access big geodata may also overcome the issue of data scarcity.

This study reveals a pivotal result for the relevant entities that contribute to mitigating the landslide risk. The highly exposed areas to landslides offer a profound spatial awareness for government agencies, spatial planners, emergency services, communities, and private owners. Thus, contributions to mitigation measures can be taken accordingly by the multifaceted stakeholders.

Acknowledgments

The authors express their gratitude to the Center for Geological Disaster, National Research, and Innovation Agency (BRIN), Indonesia, for supporting and facilitating our research group of "Spatial Modelling for Landslide Disaster". The authors also thank the editors and reviewers for their diligent efforts and valuable feedback during the publication process.

References

Achanta R, Susstrunk S (2017) Superpixels and Polygons Using Simple Non-iterative Clustering. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, 4895–4904. https://doi.org/10.1109/CVPR.2017.520

Amarasinghe MP, Kulathilaka SAS, Robert DJ, Zhou A, Jayathissa HAG (2024) Risk assessment and management of rainfall-induced landslides in tropical regions: a review. Natural Hazards 120(3): 2179–2231. https://doi.org/10.1007/s11069-023-06277-3

Berov B, Frantzova A, Ivanov P, Dobrev N, Krastanov M, Nankin R (2024) Risk assessment of landslides: Low probability scenario for the town of Kavarna, northern Black Sea coast of Bulgaria. Journal of the Bulgarian Geographical Society 50: 129–148. https://doi.org/10.3897/jbgs.e116695

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1): 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Queiroz Feitosa R, Van Der Meer F, Van Der Werff H, Van Coillie F, Tiede D (2014) Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87: 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

BNPB (2025) Data Informasi Bencana Indonesia [Indonesia Disaster Database]. https://dibi.bnpb.go.id/superset/dashboard/2/ [Accessed on 02.10.2025]

Bondarenko M, Kerr D, Sorichetta A, Tatem A (2020) Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 183 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs. https://doi.org/10.5258/SOTON/WP00685

BPS [Central Statistics Agency] (2022a) Bogor Regency in Figures 2022.

BPS [Central Statistics Agency] (2022b) Bogor City in Figures 2022.

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

Caleca F, Scaini C, Frodella W, Tofani V (2024) Regional-scale landslide risk assessment in Central Asia. Natural Hazards and Earth System Sciences 24(1): 13–27. https://doi.org/10.5194/nhess-24-13-2024

Caleca F, Tofani V, Segoni S, Raspini F, Rosi A, Natali M, Catani F, Casagli N (2022) A methodological approach of QRA for slow-moving landslides at a regional scale. Landslides 19(7): 1539–1561. https://doi.org/10.1007/s10346-022-01875-x

Corominas J, Van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2014) Recommendations for the quantitative analysis of landslide risk. Bulletin of Engineering Geology and the Environment. https://doi.org/10.1007/s10064-013-0538-8

Dash P, Sanders SL, Parajuli P, Ouyang Y (2023) Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data in an Agricultural Watershed. Remote Sensing 15(16): 4020. https://doi.org/10.3390/rs15164020

De La Cruz Huayanay A, Bazán JL, Russo CM (2024) Performance of evaluation metrics for classification in imbalanced data. Computational Statistics. https://doi.org/10.1007/s00180-024-01539-5

Di Napoli M, Miele P, Guerriero L, Annibali Corona M, Calcaterra D, Ramondini M, Sellers C, Di Martire D (2023) Multitemporal relative landslide exposure and risk analysis for the sustainable development of rapidly growing cities. Landslides 20(9): 1781–1795. https://doi.org/10.1007/s10346-023-02065-z

ESA [European Space Agency] (2015) Sentinel-2 User Handbook. ESA Standard Document, 24/07/2015 Issue 1 Rev 2.

Estoque RC, Murayama Y (2015) Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices. Ecological Indicators 56: 205–217. https://doi.org/10.1016/j.ecolind.2015.03.037

Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sensing of Environment 80(1): 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Friedman JH (2002) Stochastic gradient boosting. Computational Statistics & Data Analysis 38(4): 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Galderisi A, Limongi G (2021) A Comprehensive Assessment of Exposure and Vulnerabilities in Multi-Hazard Urban Environments: A Key Tool for Risk-Informed Planning Strategies. Sustainability 13(16): 9055. https://doi.org/10.3390/su13169055

Garcia RAC, Oliveira SC, Zêzere JL (2016) Assessing population exposure for landslide risk analysis using dasymetric cartography. Natural Hazards and Earth System Sciences 16(12): 2769-2782. https://doi.org/10.5194/nhess-16-2769-2016

Gaughan AE, Stevens FR, Linard C, Jia P, Tatem AJ (2013) High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. PLoS ONE 8: e55882. https://doi.org/10.1371/journal.pone.0055882

Gitelson AA, Merzlyak MN (1996) Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. Journal of Plant Physiology 148(3–4): 494–500. https://doi.org/10.1016/S0176-1617(96)80284-7

Goodchild MF (2018) 1.03 - Big Geodata. In: Comprehensive Geographic Information Systems. Elsevier, 19–25. https://doi.org/10.1016/B978-0-12-409548-9.09595-6

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202: 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Khan S, Bhardwaj A, Sakthivel M (2024) Accuracy Assessment of Land Use Land Cover Classification Using Machine Learning Classifiers in Google Earth Engine; A Case Study of Jammu District. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4–2024: 263–268. https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-263-2024

Kjekstad O, Highland L (2009) Economic and Social Impacts of Landslides. In: Sassa K, Canuti P (Eds), Landslides – Disaster Risk Reduction. Springer Berlin Heidelberg, Berlin, Heidelberg, 573–587. https://doi.org/10.1007/978-3-540-69970-5_30

Lillesand T, Kiefer RW, Chipman J (2015) Remote sensing and image interpretation. 7th Edition. John Wiley & Sons, Inc., Hoboken, 768 pp.

Lloyd CT, Chamberlain H, Kerr D, Yetman G, Pistolesi L, Stevens FR, Gaughan AE, Nieves JJ, Hornby G, MacManus K, Sinha P, Bondarenko M, Sorichetta A, Tatem AJ (2019)

Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. Big Earth Data 3(2): 108–139. https://doi.org/10.1080/20964471.2019.1625151

Luu C, Ha H, Thong Tran X, Ha Vu T, Duy Bui Q (2024) Landslide susceptibility and building exposure assessment using machine learning models and geospatial analysis techniques. Advances in Space Research 74(11): 5489–5513. https://doi.org/10.1016/j.asr.2024.08.046

Mallick J, Alqadhi S, Talukdar S, AlSubih M, Ahmed Mohd, Khan RA, Kahla NB, Abutayeh SM (2021) Risk Assessment of Resources Exposed to Rainfall Induced Landslide with the Development of GIS and RS Based Ensemble Metaheuristic Machine Learning Algorithms. Sustainability 13: 457. https://doi.org/10.3390/su13020457

Marpu PR, Neubert M, Herold H, Niemeyer I (2010) Enhanced evaluation of image segmentation results. Journal of Spatial Science 55(1): 55–68. https://doi.org/10.1080/14498596.2010.487850

Mateos RM, López-Vinielles J, Poyiadji E, Tsagkas D, Sheehy M, Hadjicharalambous K, Liscák P, Podolski L, Laskowicz I, Iadanza C, Gauert C, Todorović S, Auflič MJ, Maftei R, Hermanns RL, Kociu A, Sandić C, Mauter R, Sarro R, Béjar M, Herrera G (2020) Integration of landslide hazard into urban planning across Europe. Landscape and Urban Planning 196: 103740. https://doi.org/10.1016/j.landurbplan.2019.103740

McFeeters SK (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing 17(7): 1425–1432. https://doi.org/10.1080/01431169608948714

Melati DN, Astisiasari, Trinugroho (2022) An Assessment of Object-based Classification Compared to Pixel-based Classification in Google Earth Engine Using Random Forest. In: 2022 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS). IEEE, Surabaya, Indonesia, 73–78. https://doi.org/10.1109/AGERS56232.2022.10093267

Melati DN, Umbara RP, Astisiasari A, Wisyanto W, Trisnafiah S, Trinugroho T, Prawiradisastra F, Arifianti Y, Ramdhani TI, Arifin S, Anggreainy MS (2024) A comparative evaluation of landslide susceptibility mapping using machine learning-based methods in Bogor area of Indonesia. Environmental Earth Sciences 83(3): 86. https://doi.org/10.1007/s12665-023-11402-3

Muhamad N, Arshad SHM, Pereira JJ (2021) Exposure Elements in Disaster Databases and Availability for Local Scale Application: Case Study of Kuala Lumpur, Malaysia. Frontiers in Earth Science 9: 616246. https://doi.org/10.3389/feart.2021.616246

Müller AC, Guido S (2016) Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc., Sebastopol, 392 pp.

Nasiri V, Deljouei A, Moradi F, Sadeghi SMM, Borz SA (2022) Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. Remote Sensing 14(9): 1977. https://doi.org/10.3390/rs14091977

Ouma YO, Keitsile A, Nkwae B, Odirile P, Moalafhi D, Qi J (2023) Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. European Journal of Remote Sensing 56(1): 2173659. https://doi.org/10.1080/22797254.2023.2173659

Ozturk U, Bozzolan E, Holcombe EA, Shukla R, Pianosi F, Wagener T (2022) How climate change and unplanned urban sprawl bring more landslides. Nature 608(7922): 262–265. https://doi.org/10.1038/d41586-022-02141-9

Quesada-Román A, Campos-Durán D (2023) Natural Disaster Risk Inequalities in Central America. Papers in Applied Geography 9(1): 36-48. https://doi.org/10.1080/23754931.2022.2081814

Rahman G, Bacha AS, Ul Moazzam MF, Rahman AU, Mahmood S, Almohamad H, Al Dughairi AA, Al-Mutiry M, Alrasheedi M, Abdo HG (2022) Assessment of landslide susceptibility, exposure, vulnerability, and risk in shahpur valley, eastern hindu kush. Frontiers in Earth Science 10: 953627. https://doi.org/10.3389/feart.2022.953627

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth-Science Reviews 180: 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Riaz MT, Basharat M, Brunetti MT, Riaz MT (2023) Semi-quantitative landslide risk assessment of district Muzaffarabad, northwestern Himalayas, Pakistan. Stochastic Environmental Research and Risk Assessment 37(9): 3551–3570. https://doi.org/10.1007/s00477-023-02462-9

Sahin EK (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences 2(7): 1308. https://doi.org/10.1007/s42452-020-3060-1

Sarafova E (2022) Data quality assessment of Copernicus Climate Change Service health domain data for the development of disaster risk reduction plans. Journal of the Bulgarian Geographical Society 46: 13–23. https://doi.org/10.3897/jbgs.e85567

Sarafova E (2023) Communicating disasters to children through digital learning activities, geospatial data and platforms. Journal of the Bulgarian Geographical Society 48: 73–84. https://doi.org/10.3897/jbgs.e106818

Silalahi FES, Pamela, Arifianti Y, Hidayat F (2019) Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. Geoscience Letters 6(1): 10. https://doi.org/10.1186/s40562-019-0140-4

Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62(1): 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE 10(2): e0107042. https://doi.org/10.1371/journal.pone.0107042

Trisnafiah S, Trinugroho T, Astisiasari A, Riyandari R, Handayani MP, Melati DN, Umbara RP, Arifianti Y, Ramdhani TI, Sehah S (2024) Spatial-Based Landslide Vulnerability Index Assessment in Bogor Area, Indonesia. Geosfera Indonesia 9: 16. https://doi.org/10.19184/geosi.v9i1.43643

Umbara RP, Melati DN, Astisiasari, Wisyanto, Trisnafiah S, Trinugroho, Arifianti Y, Prawiradisastra F, Ramdhani TI, Arifin S, Anggreainy MS (2024) Utilization of Frequency Ratio and Logistic Regression Model for Landslide Susceptibility Mapping in Bogor Area. International Journal on Advanced Science, Engineering and Information Technology 14(2): 528–539. https://doi.org/10.18517/ijaseit.14.2.19345

UNISDR [United Nations Office for Disaster Risk Reduction] (2017) National Disaster Risk Assessment. UNISDR, Geneva.

Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Engineering Geology 102(3–4): 112–131. https://doi.org/10.1016/j.enggeo.2008.03.010

Wicaksono YS, Sihombing FMH, Indra TL (2020) Landslide susceptibility map of Bogor Area using analytical hierarchy process. IOP Conference Series: Earth and Environmental Science 538(1): 012050. https://doi.org/10.1088/1755-1315/538/1/012050

WorldPop (2025) WorldPop Hub Population Counts https://hub.worldpop.org/project/categories?id=3 [Accessed on 13.01.2025]

Zhang Z, Li Z, Li Z (2025) Evaluating the change and trend of construction land in Changsha City based GeoSOS-FLUS model and machine learning methods. Scientific Reports 15(1): 9602. https://doi.org/10.1038/s41598-025-93689-9

Additional information

Conflict of interest

No conflict of interest was declared.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

This study received funding from the Indonesian National Research and Innovation Agency (BRIN), under the Program of Research and Innovation Technology for Disaster and Earth Resources (SK No.1/III.4/HK/2025) and Program of Research and Innovation Technology for BIGDATA (SK No. 1/III.6/HK/2025).

Author contributions

Conceptualization: AA, DNM, WW. Data curation: WW, SS, AA, DNM, LYA. Formal analysis: WW, AA, DNM. Funding acquisition: DNM, RPU, TIR, YA. Investigation: DNM, AA, WW, LYA, SS. Methodology: WW, DNM, AA. Project administration: RPU, TIR, YA, DNM. Resources: WW, LYA, AA, SS, DNM. Software: WW, AA, DNM. Supervision: AA, DNM, WW. Validation: WW, DNM, AA. Visualization: TIR, AA, WW, DNM, SS, YA, TT. Writing - original draft: DNM, AA, LYA, WW, RPU, SS. Writing - review and editing: SS, YA, AA, TT, DNM, WW, TIR.

Author ORCIDs

Astisiasari Astisiasari https://orcid.org/0000-0003-1618-3000
Wisyanto Wisyanto https://orcid.org/0009-0007-3583-6536
Dian Melati https://orcid.org/0009-0002-0672-6980
Sukristiyanti Sukristiyanti https://orcid.org/0000-0002-1945-1878
Raditya Umbara https://orcid.org/0009-0003-7386-3157
Yukni Arifianti https://orcid.org/0000-0003-3838-2529
Trinugroho Trinugroho https://orcid.org/0000-0002-0419-2922
Lian Andikasari https://orcid.org/0009-0007-3969-9348
Taufik Ramdhani https://orcid.org/0000-0001-7436-4884

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.