Assessing Landslide Vulnerability in Deforested Areas of Sumatra Island Using Remote Sensing and Machine Learning

Ifandra Kusuma Cahya Reynaldi[1], Anjar Dimara Sakti[2], Deni Suwardhi[3], Agung Budi Harto[4]
{ifandrakusumacahyareynaldi@gmail.com[1], anjar@itb.ac.id[2], deni.suwardhi@itb.ac.id[3]}

Remote Sensing and Geographic Information Science Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia[1,2,3,4], Center for Remote Sensing, Institut Teknologi Bandung, Bandung 40132, Indonesia[1,2,3,4]

Abstract. Sumatra is among 11 regions in the world that significantly contribute to global deforestation. In the period 2001-2020, deforestation in Sumatra has reached more than three million hectares. This has made hydrometeorological disasters a reality in Sumatra due to the loss of rainwater catchment. One of them is landslides that often occur and have endangered residents and productive areas in Sumatra. The development of ecosystem-based mitigation can be done to reduce risks and losses that may occur. However, it is necessary to conduct preliminary studies to assess the likelihood of future hazards. This can be done by utilizing remote sensing technology and geographic information systems using various geospatial data. Supported by a machine learning approach, it can improve the quality of hazard assessment. It was found that Sumatera has landslide hazard vulnerability dominated from medium to high level in mountain area, but not in deforested area.

Keywords: landslide, deforestation, remote sensing, machine learning, hazard.

1 Introduction

Landslides are a natural disaster that occur in mountainous regions all over the world, causing thousands of fatalities each year. Deforestation has been found to increase landslide occurrence[1], [2], [3]. Deforestation can destabilizes the soil as tree roots decay, further increasing landslide hazard, especially rainfall-induced landslides [2], [4]. Forests and trees are useful in landslide reduction, and landslides are a growing hazard [5]. It is important to recognize the link between deforestation and landslides and take measures to prevent deforestation and promote reforestation in landslide-prone areas.

Sumatra is among 11 regions in the world that contribute to 80% of global deforestation [6]. In the period 2001-2020, deforestation in Sumatra has reached more than three million hectares [6]. This has made hydrometeorological disasters a reality in Sumatra due to the loss of rainwater catchment [7], [8], [9]. One of them is landslides that often occur and have

RESILIENCE 2023, November 24, Jakarta, Indonesia

DOI 10.4108/eai.24-11-2023.2346387

endangered residents and productive areas in Sumatra. These two events are interrelated and will influence each other's impact.

However, it is necessary to conduct preliminary studies to assess the likelihood of future hazards. This is possible by utilizing remote sensing and geographic information systems (GIS) to determine potential hazards [10]. Then, supported by the approach of various available machine learning methods, such as decision trees, artificial neural networks, random forests, and support vector machines, which have the advantages of improving prediction accuracy and reducing measurement errors, it will be possible to improve the quality of hazard assessments [11], [12], [13].

This study aims to identify landslide vulnerability using machine learning algorithms, then link it to the phenomenon of deforestation on the island of Sumatra. Hopefully, this initial study can serve as an initial trigger in the development of ecosystem-based adaptation and solutions to deforestation and landslides.

2 Methodology

The location of this landslide susceptibility mapping study is Sumatera Island. The study area used is shown in **Figure 1**.

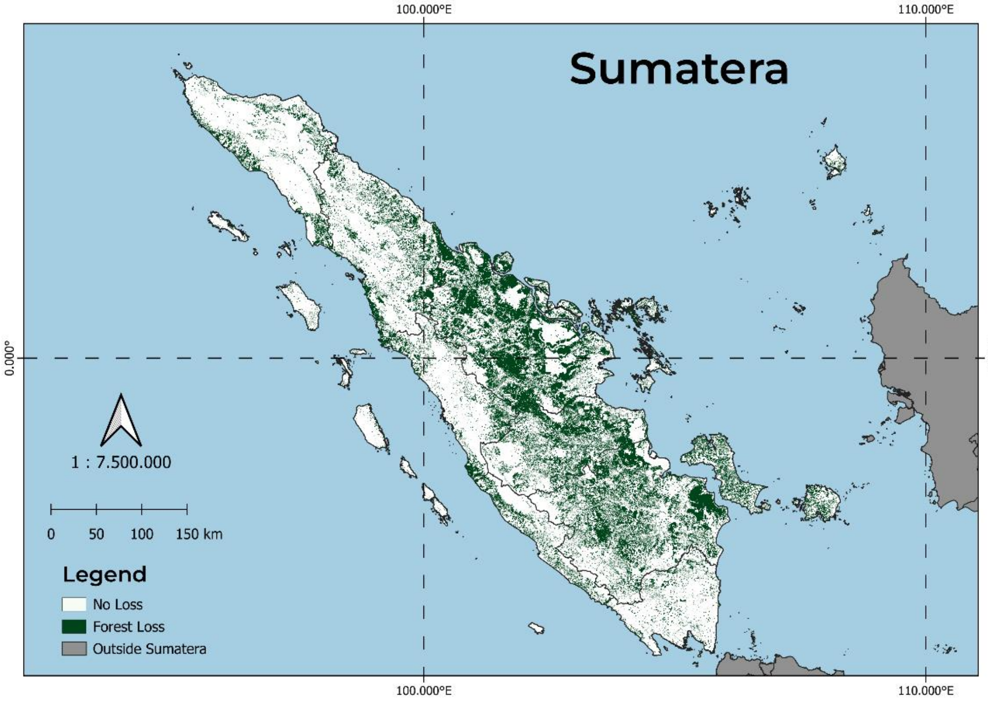

Fig. 1. Study Area.

In modeling landslide susceptibility using machine learning, training data and prediction parameters are required. The training data used in the modeling is historical point data of landslide events obtained from NASA. On the other hand, the prediction parameters used are elevation data, slope data, soil water content data, normalized difference vegetation index (NDVI) data, enhanced vegetation index (EVI) data, precipitation data, and land cover data. The data used is shown in Table 1.

Table 1: Data.

No	Data	Product	Data Type / Resolution	References
1	Landslide repository events points	NASA Cooperative Open Online Landslide Repository (COOLR)	Tabular	[14], [15]
2	Administrative boundaries	RBI Map	Vector	[16]
3	Elevation	FABDEM (Forest and Buildings removed Copernicus 30m DEM)	Raster / 30m	[17]
4	Slope	FABDEM (Forest and Buildings removed Copernicus 30m DEM)	Raster / 30m	[17]
5	Soil water content	OpenLandMap Soil Water Content	Raster / 250m	[18]
6	Normalized difference vegetation index (NDVI)	MOD13Q1.006 Terra Vegetation Indices 16-Day Global	Raster / 250m	[19]
7	Enhanced vegetation index (EVI)	MOD13Q1.006 Terra Vegetation Indices 16-Day Global	Raster / 250m	[19]
8	Precipitation	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho	Raster / 5km	[20]
9	Land Cover	Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3	Raster / 100m	[21]
10	Forest Loss	Hansen Global Forest Change v1.10	Raster / 30m	[22]

In this study, the process is generally divided into 2 main parts: landslide susceptibility modeling using machine learning and landslide susceptibility modeling in deforestation areas. The methodology used is shown in **Figure 2**.

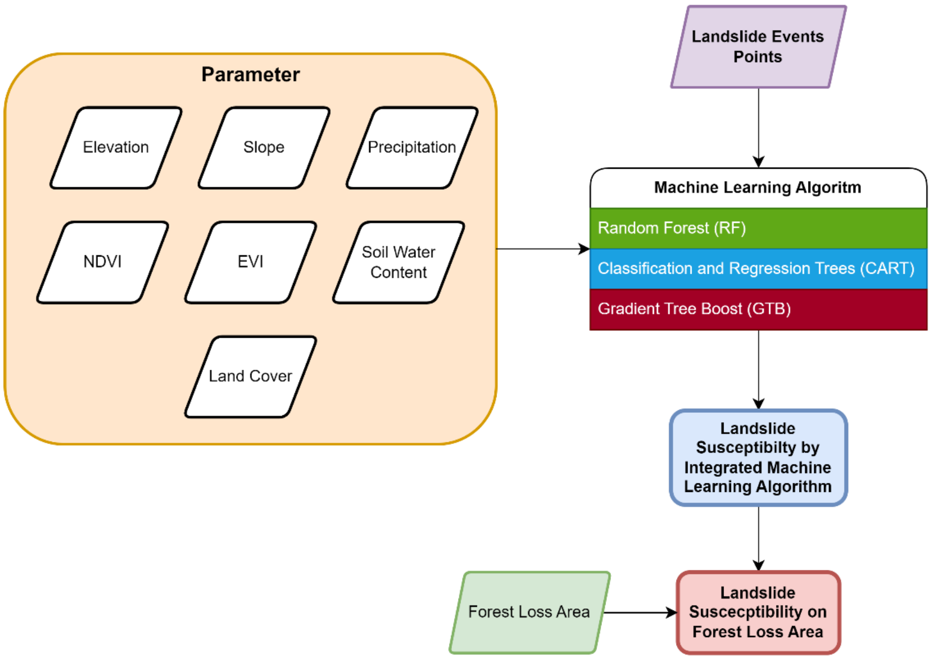

Fig. 2. Methodology.

Landslide susceptibility modeling is conducted using supervised classification method using machine learning algorithm based on landslide event points and predictor parameters. The predictor parameters are used as parameters that are taken into account in assessing the probability of landslide susceptibility of other areas based on the value of the predictor factor at the landslide event points. Modeling was conducted using 3 (three) machine learning algorithms, namely Random Forest (RF), Classification and Regression Trees (CART), and Gradient Tree Boost (GTB).

Random forest is a supervised classification algorithm, and is an ensemble method using a decision tree model so that each tree corresponds to an independently sampled subset of data using a bootstrap technique [23]. Random selection of predictor variables is used to divide each node of the developed tree to minimize classification error. Classification and Regression Tree (CART) is a rule-based algorithm that generates binary trees through "binary recursive partitioning", a process that splits nodes into yes/no answers as predictor values [24]. If the dependent variable is a categorical scale, CART will produce a classification tree, and if the dependent variable is continuous data, then CART will produce a regression tree [25]. Gradient Tree Boost (GTB) algorithm is a combination of decision tree and boosting algorithm proposed by Friedman in 2001 [26]. Boosting refers to the combination of multiple weak classifiers to achieve a strong classifier, and gradient refers to increasing flexibility and ease when the model minimizes the loss function [27], [28].

The development of landslide susceptibility of multi-machine learning algorithm is done by combining the three machine learning algorithm models to form an agreement index value of landslide susceptibility of an area. The landslide susceptibility model of each algorithm has a susceptibility index value of 0 (zero) to 1 (one) which indicates a non-susceptible to susceptible area. A summation of the susceptibility index values is done to produce an agreement index value in the range of 0 (zero) to 3 (three). This is done to increase sensitivity, which is the ability of the model to detect a class correctly [29]. Landslide susceptibility index based on multi-machine learning algorithm is formed based on Equation 1.

$$LSI = RF + CART + GTB \tag{1}$$

Furthermore, to form the landslide susceptibility in deforested areas, the resulting landslide susceptibility is clipped based on the deforestation area. After that, the landslide susceptibility of deforested area in Sumatera Island is obtained.

3 Results

3.1 Landslide Susceptibility Based on Machine Learning Algorithm

The landslide susceptibility model for each machine learning algorithm is shown in **Figure 3.**

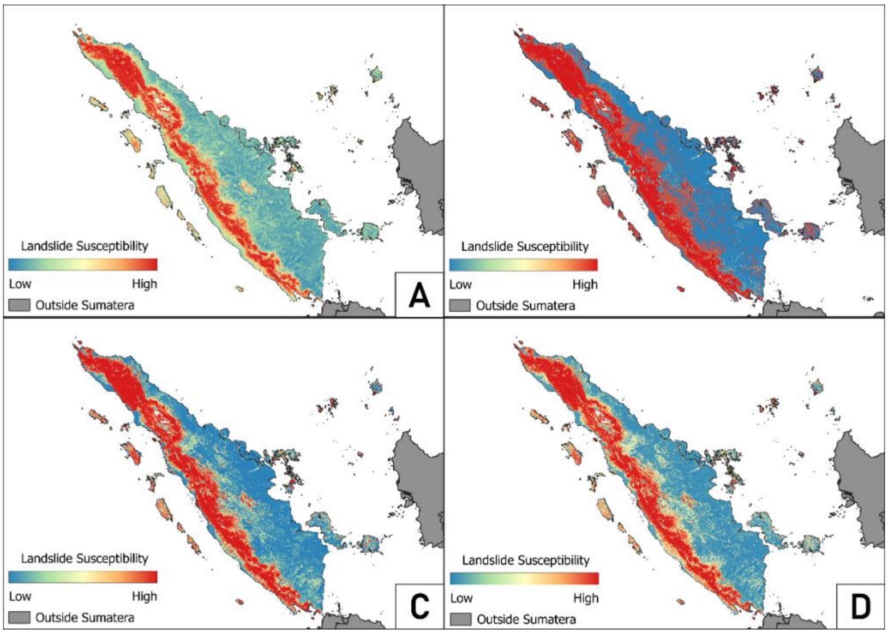

Fig. 3. Landslide susceptibility based on machine learning algorithm: (A) random forest algorithm; (B) classification and regression trees algorithm; (C) gradient tree boost algorithm; (D) multi-machine learning algorithm.

The resulting visuals show that the RF algorithm and GTB algorithm models have an even distribution of colors, when compared to the CART algorithm model. It can be seen that the CART algorithm has a fairly drastic color distribution, there are only low and high classes in the resulting model. While in the multi-machine learning algorithm model, the agreement index of the three machine learning algorithms used previously is obtained. It can be seen that all algorithms have a high level of vulnerability in mountainous areas on the island of Sumatra, or in areas with high elevation or slope.

Next, the area for each vulnerability class was calculated and the results are shown in **Figure 4.**

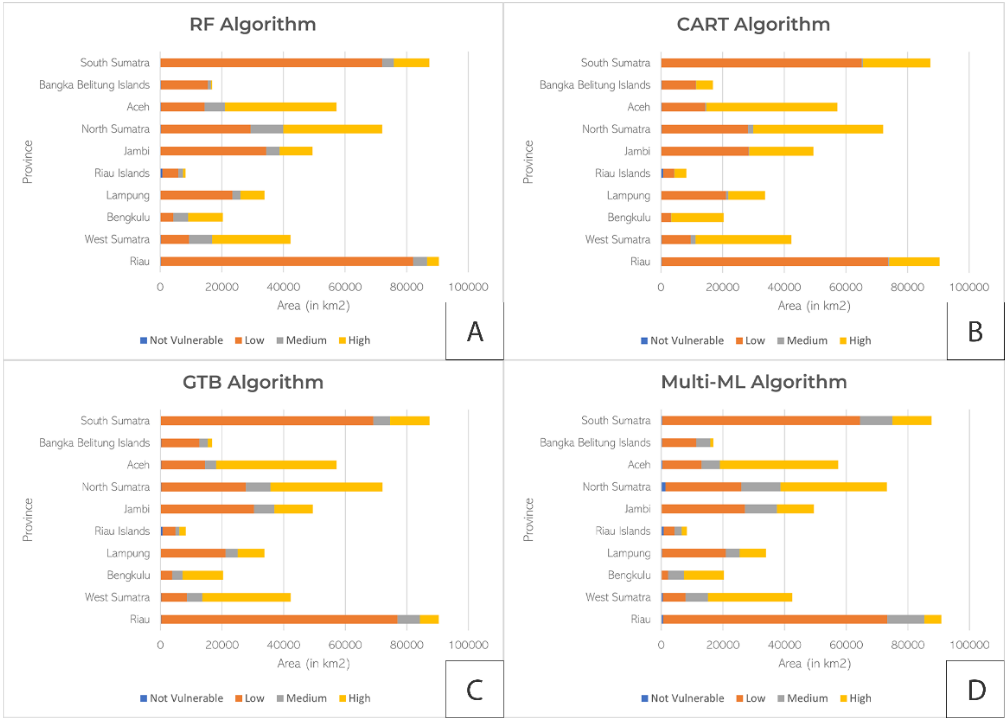

Fig. 4. Vulnerability area of each algorithm: (A) random forest algorithm; (B) classification and regression trees algorithm; (C) gradient tree boost algorithm; (D) multi-machine learning algorithm.

Based on the distribution of data, in accordance with the visualization, it can be seen that the CART algorithm model has a poor distribution of predicted data where only a few medium landslide vulnerability classes are classified. In contrast to other algorithms, which have medium vulnerability class although not dominant. In multi-machine learning algorithm, the medium vulnerability class is more evenly distributed. This indicates a better classification result.

Based on the graphical results obtained, it can be seen that for all algorithms, the area is dominated by low landslide vulnerability class compared to the whole Sumatera Island. However, it can be seen that some provinces such as Aceh Province, North Sumatra Province, Bengkulu Province, and West Sumatra Province have medium to high vulnerability levels that dominate their areas. This is because these provinces are dominated by mountainous areas with high elevation and slope, making them more vulnerable to landslides.

3.2 Landslide Susceptibility in Deforested Area

The landslide susceptibility model for each machine learning algorithm is shown in **Figure 5.**

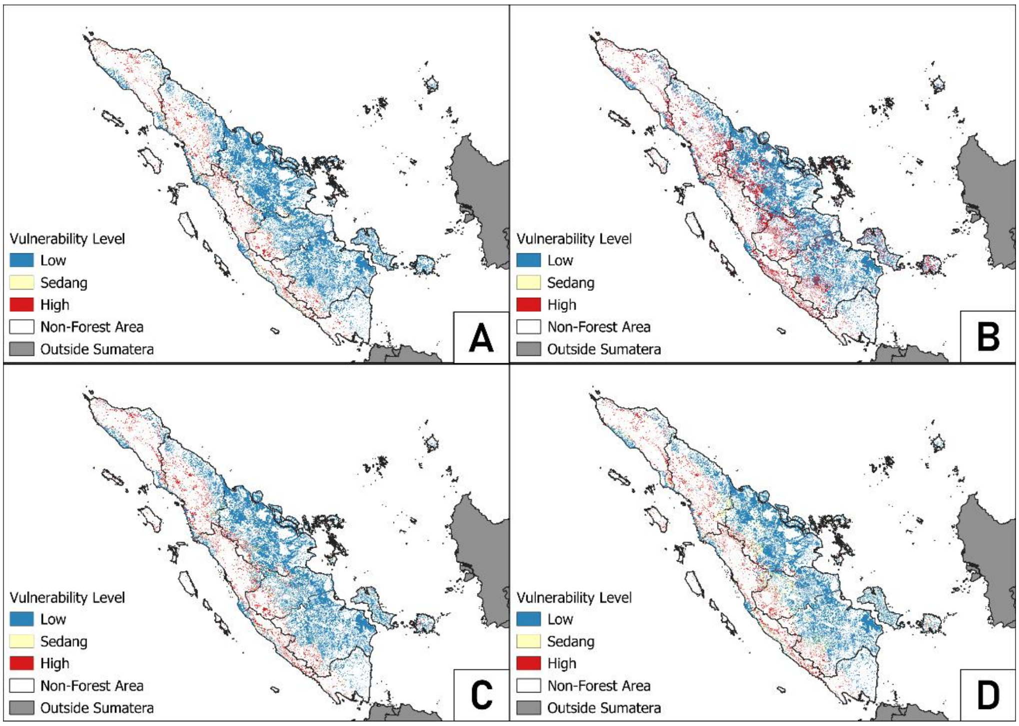

Fig. 5. Landslide susceptibility on deforestation area: (A) random forest algorithm; (B) classification and regression trees algorithm; (C) gradient tree boost algorithm; (D) multi-machine learning algorithm.

In the visual of landslide vulnerability in deforestation areas, it can be seen that landslide vulnerability in deforestation areas is drastically reduced in terms of distribution. This shows that not all areas of Sumatera Island are deforested. Areas with high elevation and slope, which are areas with high landslide susceptibility class, are less likely to experience deforestation compared to areas with medium or low vulnerability class. This means that areas with high landslide vulnerability class in Sumatera Island tend not to experience deforestation.

Next, the area for each vulnerability class was calculated and the results are shown in **Figure 6**.

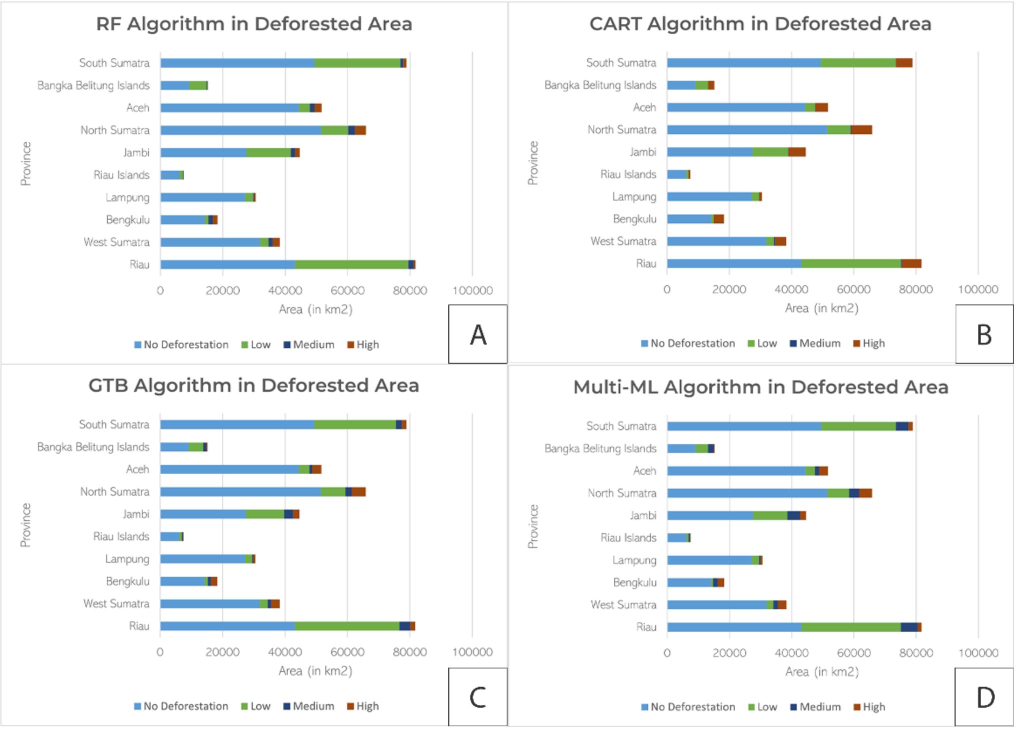

Fig. 6. Vulnerability area of each algorithm in deforested area: (A) random forest algorithm; (B) classification and regression trees algorithm; (C) gradient tree boost algorithm; (D) multi-machine learning algorithm.

Based on the results of the graph, it is found that the area for the high vulnerability class decreases drastically according to its spatial distribution. It is also found that the area without deforestation is dominant in Sumatra. Spatially, these no-deforestation areas include areas with high elevation and slope. This supports the previous argument, where deforestation tends not to occur in areas with high landslide vulnerability class that have high elevation and slope. It can also be interpreted that deforestation and reforestation tend to occur repeatedly in an area with medium to low vulnerability class that has relatively lower elevation and slope. It can be concluded that deforestation in Sumatra tends to be caused by factors other than landslides, such as land conversion, illegal logging, or land and forest fires caused by hotspots including one that occurred in Riau and West Sumatra Provinces [30], [31], [32], [33].

4 Conclusions

This study found that medium to high landslide vulnerability is dominant in mountainous areas with high elevation and slope. However, these mountainous areas tend not to experience deforestation. This means that landslides that occur in Sumatra are not caused by deforestation, and deforestation that occurs in Sumatra is not caused by landslides. It can be assumed that landslides or deforestation that occur in Sumatra are caused by other factors.

References

[1] C. W. Runyan and P. D'Odorico, "Bistable dynamics between forest removal and landslide occurrence," *Water Resour Res*, vol. 50, no. 2, pp. 1112–1130, Feb. 2014, doi: 10.1002/2013WR014819.

[2] A. Depicker *et al.*, "Historical dynamics of landslide risk from population and forest-cover changes in the Kivu Rift," *Nat Sustain*, vol. 4, no. 11, pp. 965–974, Aug. 2021, doi: 10.1038/S41893-021-00757-9).

[3] RFA's Vietnamese Service, "Deforestation in Vietnam Blamed for Recent Deadly Landslides — Radio Free Asia." Accessed: Oct. 25, 2023. [Online]. Available: https://www.rfa.org/english/news/vietnam/deforestation-11112020125229.html

[4] P. Lehmann, J. von Ruette, and D. Or, "Deforestation Effects on Rainfall-Induced Shallow Landslides: Remote Sensing and Physically-Based Modelling," *Water Resour Res*, vol. 55, no. 11, pp. 9962–9976, Nov. 2019, doi: 10.1029/2019WR025233.

[5] K. Forbes, J. Broadhead, and S. L. Kuriakose, "Forests and landslides : the role of trees and forests in the prevention of landslides and rehabilitation of landslide-affected areas in Asia.," 2011.

[6] BBC News Indonesia, "Hutan Sumatera dan Kalimantan Sumbang Deforestasi Global - BBC News Indonesia." Accessed: Oct. 25, 2023. [Online]. Available: https://www.bbc.com/indonesia/berita_indonesia/2015/04/150428_sains_hutan

[7] T. Yanuarto, "Bencana Hidrometeorologi Basah Tiga Wilayah Sumatra Barat, 2 Warga Hilang - BNPB." Accessed: Oct. 25, 2023. [Online]. Available: https://bnpb.go.id/berita/bencana-hidrometeorologi-basah-tiga-wilayah-sumatra-barat-2-warga-hilang

[8] Q. Rostanti, "Masyarakat di Sumut Diimbau Waspadai Bencana Hidrometeorologi | Republika Online." Accessed: Oct. 25, 2023. [Online]. Available: https://news.republika.co.id/berita/s2dgvl425/masyarakat-di-sumut-diimbau-waspadai-bencana-hidrometeorologi

[9] A. Amindoni and A. Adzkia, "Banjir dan bencana beruntun di tengah cuaca ekstrem, 'Menurut pemerintah itu anomali cuaca, kami menyebutnya krisis iklim' - BBC News Indonesia." Accessed: Oct. 25, 2023. [Online]. Available: https://www.bbc.com/indonesia/indonesia-56007558

[10] S. Samanta, D. K. Pal, and B. Palsamanta, "Flood susceptibility analysis through remote sensing, GIS and frequency ratio model," *Appl Water Sci*, vol. 8, no. 2, pp. 1–14, May 2018, doi: 10.1007/S13201-018-0710-1/TABLES/5.

[11] M. Bordbar, H. Aghamohammadi, H. R. Pourghasemi, and Z. Azizi, "Multi-hazard spatial modeling via ensembles of machine learning and meta-heuristic techniques," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-05364-y.

[12] S. Ghaffarian and S. Emtehani, "Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery," *Climate*, vol. 9, no. 4, Apr. 2021, doi: 10.3390/cli9040058.

[13] A. R. M. T. Islam, M. M. R. Bappi, S. Alqadhi, A. A. Bindajam, J. Mallick, and S. Talukdar, "Improvement of flood susceptibility mapping by introducing hybrid

ensemble learning algorithms and high-resolution satellite imageries," *Natural Hazards*, vol. 119, no. 1, pp. 1–37, Oct. 2023, doi: 10.1007/S11069-023-06106-7/METRICS.

[14] D. Kirschbaum, T. Stanley, and Y. Zhou, "Spatial and temporal analysis of a global landslide catalog," *Geomorphology*, vol. 249, pp. 4–15, Nov. 2015, doi: 10.1016/J.GEOMORPH.2015.03.016.

[15] D. B. Kirschbaum, R. Adler, Y. Hong, S. Hill, and A. Lerner-Lam, "A global landslide catalog for hazard applications: Method, results, and limitations," *Natural Hazards*, vol. 52, no. 3, pp. 561–575, Jan. 2010, doi: 10.1007/S11069-009-9401-4.

[16] Badan Informasi Geospasial Republik Indonesia, "Peta Rupabumi Digital Indonesia." Accessed: Oct. 25, 2023. [Online]. Available: http://tanahair.indonesia.go.id/portal-web/

[17] L. Hawker *et al.*, "A 30 m global map of elevation with forests and buildings removed," *Environmental Research Letters*, vol. 17, no. 2, Feb. 2022, doi: 10.1088/1748-9326/AC4D4F.

[18] T. Hengl and S. Gupta, "Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution," Apr. 2019, doi: 10.5281/ZENODO.2784001.

[19] K. Didan, "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]," *NASA EOSDIS Land Processes DAAC*, 2015, doi: 10.5067/MODIS/MOD13Q1.006.

[20] J. T. Abatzoglou, S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch, "TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015," *Scientific Data 2018 5:1*, vol. 5, no. 1, pp. 1–12, Jan. 2018, doi: 10.1038/sdata.2017.191.

[21] M. Buchhorn, M. Lesiv, N. E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus Global Land Cover Layers—Collection 2," *Remote Sensing 2020, Vol. 12, Page 1044*, vol. 12, no. 6, p. 1044, Mar. 2020, doi: 10.3390/RS12061044.

[22] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science (1979)*, vol. 342, no. 6160, pp. 850–853, Nov. 2013, doi: 10.1126/SCIENCE.1244693/SUPPL_FILE/HANSEN.SM.PDF.

[23] L. Breiman, M. Last, and J. Rice, "Random Forests: Finding Quasars," *Statistical Challenges in Astronomy*, pp. 243–254, May 2003, doi: 10.1007/0-387-21529-8_16.

[24] A. M. Youssef, H. R. Pourghasemi, Z. S. Pourtaghi, and M. M. Al-Katheeri, "Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia," *Landslides*, vol. 13, no. 5, pp. 839–856, Oct. 2016, doi: 10.1007/S10346-015-0614-1.

[25] P. Subarkah, A. N. Ikhsan, and A. Setyanto, "The effect of the number of attributes on the selection of study program using classification and regression trees algorithms," *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, pp. 1–5, Jul. 2018, doi: 10.1109/ICITISEE.2018.8721030.

[26] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *https://doi.org/10.1214/aos/1013203451*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/AOS/1013203451.

[27] Z. Ye *et al.*, "Using Machine Learning Algorithms Based on GF-6 and Google Earth Engine to Predict and Map the Spatial Distribution of Soil Organic Matter Content,"

Sustainability 2021, Vol. 13, Page 14055, vol. 13, no. 24, p. 14055, Dec. 2021, doi: 10.3390/SU132414055.

[28] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning", Accessed: Apr. 30, 2023. [Online]. Available: http://creativecommons.org/publicdomain/zero/1.0/

[29] V. Kotu and B. Deshpande, "Data Mining Process," *Predictive Analytics and Data Mining*, pp. 17–36, 2015, doi: 10.1016/B978-0-12-801460-8.00002-1.

[30] A. D. Sakti *et al.*, "Spatial Prioritization for Wildfire Mitigation by Integrating Heterogeneous Spatial Data: A New Multi-Dimensional Approach for Tropical Rainforests," *Volume 14, Issue 3*, vol. 14, no. 3, Jan. 2022, doi: 10.3390/rs14030543.

[31] A. S. Thoha *et al.*, "Spatial distribution of 2019 forest and land fires in Indonesia," *J Phys Conf Ser*, vol. 2421, no. 1, Jan. 2023, doi: 10.1088/1742-6596/2421/1/012035.

[32] E. Frimawaty, "Mapping data on Indonesia's worst forest and land fires of palm oil cultivation lands," *E3S Web of Conferences*, vol. 211, Jan. 2020, doi: 10.1051/e3sconf/202021105002.

[33] T. Fanin and G. R. Van Der Werf, "Relationships between burned area, forest cover loss, and land cover change in the Brazilian Amazon based on satellite data," *Biogeosciences*, vol. 12, no. 20, pp. 6033–6043, Oct. 2015, doi: 10.5194/BG-12-6033-2015.