GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 18 • Number 2 • 2024

https://doi.org/10.7494/geom.2024.18.2.73

Hana Listi Fitriana[1], Rido Dwi Ismanto[2], Jessica Stephanie Tulus[3], Atriyon Julzarika[4], Jalu Tejo Nugroho[5], Johannes Manalu[6]

Comparison of Statistical and Machine-Learning Model for Analyzing Landslide Susceptibility in Sumedang Area, Indonesia

Abstract: Landslides have produced several recurrent dangers, including losses of life and property, losses of agricultural land, erosion, population relocation, and others. Landslide mitigation is critical since population and economic expansion are rapidly followed by significant infrastructure development, increasing the risk of catastrophes. At an early stage in landslide-disaster mitigation, landslide-risk mapping must give critical information to help policies limit the potential for landslide damage. This study will utilize the comparative frequency ratio (FR) and random forest (RF) techniques; they will be utilized to properly investigate the distribution of flood vulnerability in the Sumedang area. This study has identified 12 criteria for developing a landslide-susceptibility model in the research region based on the features of past disasters in the research area. The FR and RF models scored 88 and 81% of the AUC value, respectively. Based on the McNemar test, the FR and RF models featured the same performance in determining the landslide-vulnerability level performances in Sumedang. They performed well in assessing landslides in the research region; therefore, they may be used as references in landslide prevention and references in future regional development plans by the stakeholders.

Keywords: landslide, susceptibility analysis, frequency ratio, random forest, Sumedang

Received: December 8, 2022; accepted: February 6, 2024

[1] Department of Water Resources Management, West Java Province, Indonesia, email: hanalisti@yahoo.com (corresponding author), https://orcid.org/0000-0001-6752-8414
[2] National Research and Innovation Agency (BRIN), Research Center for Computing, Cibinong, Indonesia, email: rido001@brin.go.id, https://orcid.org/0000-0002-2903-6731
[3] University of Indonesia, Department of Geophysics, Depok, Indonesia, email: jessica.stephanie81@sci.ui.ac.id, https://orcid.org/0000-0002-0619-1667
[4] National Research and Innovation Agency (BRIN), Research Center for Limnology and Water Resources, Cibinong, Indonesia, email: atriyon.julzarika@brin.go.id, https://orcid.org/0000-0002-4531-5327
[5] National Research and Innovation Agency (BRIN), Research Center for Geoinformatics, Jakarta, Indonesia, email: jalu001@brin.go.id, https://orcid.org/0000-0002-8615-2640
[6] National Research and Innovation Agency (BRIN), Research Center for Geoinformatics, Jakarta, Indonesia, email: rjoh001@brin.go.id, https://orcid.org/0009-0000-1761-3725

1. Introduction

A landslide is a naturally occurring movement of soil or rock mass on a slope that is brought on by rainfall, an earthquake, or interference from a slope, among other things [1]. In addition to other elements like slope gradient, ground saturation, and land cover, the geological characteristics like the rock type, structure, soil type, and weathering depth also affect the likelihood of landslides [2]. This phenomenon has also happened in many other places in the world, and it has caused environmental harm and land degradation in addition to losses in terms of people and money [3]. Moreover, landslides are terrible natural disasters every year in mountainous regions across the world that result in significant injuries to people, fatalities, and significant property destruction [4]. Although the physical causes of landslides cannot be eliminated through geological examinations, engineering, nor land-use management for reducing landslide hazards, it is essential to understand their potential exposure. The literature generally agrees that identifying possible landslide-prone regions, looking into landslide patterns and how they interact with post-emergency responses, hazard mitigation, and risk prevention are all essential [5]. For instance, the Indonesian government and research institutions in the field have carried out mitigation measures such as protecting landslide-prone areas, installing early warning systems that are based on the monitoring of soil conditions, installing structures such as piles and retaining walls, and so on in order to minimize losses. Therefore, in order to support policies in development to reduce the potential for landslide damage, landslide-vulnerability mapping must provide vital information as an early stage in landslide-disaster mitigation. This mitigation of landslide-prone hazards is made to help to reduce the risks to humans and property and make humans more familiar with and adapt to landslide-hazard-mitigation procedures.

During the last decade, various techniques and approaches have been developed to study landslide susceptibility, including qualitative [6], statistical [7–10], numerical [1, 11], and through the use of machine learning [12–16]. Machine learning is now becoming more widely used in landslide prevention, as it can provide optimal, accurate, efficient, and effective results with proper conditioning. As a result, it is possible to develop a reliable landslide-susceptibility model [12, 13, 17–19]. Researchers worldwide have also developed various research combinations in landslide-susceptibility research. For example, some researchers have assessed the efficacy of commonly used hybrid multi-criteria decision-making (MCDM) models for the production of landslide-susceptibility maps (LSMs) in the Alamut watershed in Iran. Those models including the integrated index method (IIM), AHP-TOPSIS (the analytic hierarchy process [AHP] combined with the technique for order of preference by similarity to the ideal solution [TOPSIS]), and AHP-VIKOR (AHP combined with viekriterijumsko kompromisno rangiranje [VIKOR]) [3]. Moreover, a geographical information system was used to conduct

a landslide-susceptibility assessment at Canada Hill, Sarawak, Malaysia, by utilizing mixed bivariate statistics and expert consultation techniques [2]. Other research has discussed a hybrid intelligent strategy for landslide-susceptibility mapping in the Bijar area of Kurdistan Province (Iran) based on a naive Bayes trees (NBT) and random subspace (RS) ensemble [20]. Generally, the success of a method depends on the geographical nature of each area that is studied [15]. Therefore, developing and comparing existing methods in determining landslide susceptibility in a research area are not impossible.

This study uses various factors from previous studies as inputs in this research model; these factors include geology [2], lithology [21], soil type [17], elevation [22], slope [21], rainfall [3], land use/land cover (LULC) [20], distance to lineament [22], distance to river [20], and road distance [20]. Our research area is an area that is prone to earthquakes, as it is situated near the Cileunyi-Tanjungsari, Kendeng and Baribis fault lines. Since it has active ground movement activity, this research explicitly adds the land movement vulnerable zone (ZKGT) parameter that was created by the Indonesian Ministry of Energy and Mineral Resources. In addition, 51.68% of the topography of the research area undulates in a hilly area, so it is necessary to add a parameter of the topographic wetness index (TWI) according to [23], because topography is an essential control in hydrological spatial distribution. Furthermore, the landslide-susceptibility analysis in this study will use statistical and machine-learning methods; namely, frequency ratio (FR) and random forest (RF), respectively. This is based on the fact that FR had enjoyed good results in landslide statistical models [8, 22, 24–29]. In addition, RF has also resulted in good performance in landslide susceptibility according to [5, 12, 14, 21, 30–34]. To evaluate the performance of the two approaches that are discussed above, receiver operating characteristic (ROC) curves and the accompanying area under the ROC curve (AUC) will be employed. A higher AUC value suggests that the model performs better. Furthermore, the McNemar test will determine whether or not both strategies work equally well. Therefore, this study aims to compare statistical methods and machine learning to be used as references in landslide prevention and development planning.

2. Material and Methods

2.1. Study Area

Sumedang Regency (Fig. 1) is one of the areas in West Java Province that consists of 26 sub-districts and 270 villages; the province has a population of 1,152,507 and a population density of 739/km^2 according to Central Bureau of Statistics (BPS) data from 2021. This regency is situated at a latitude position of 6°35'S–7°02'S and a longitude of 107°44'E–108°13'E and is located 48 km from Bandung (the capital

city of West Java Province). It can also be noticed that Sumedang is a hilly and mountainous area, with altitudes of 18 to 1996 meters above sea level. Most of the Sumedang area is mountainous, with a small portion of lowlands in the northern region. Geologically, this study region is located along the Cileunyi-Tanjungsari and Baribis fault lines, having active geological formations that cause the soil to be unstable and shift readily.

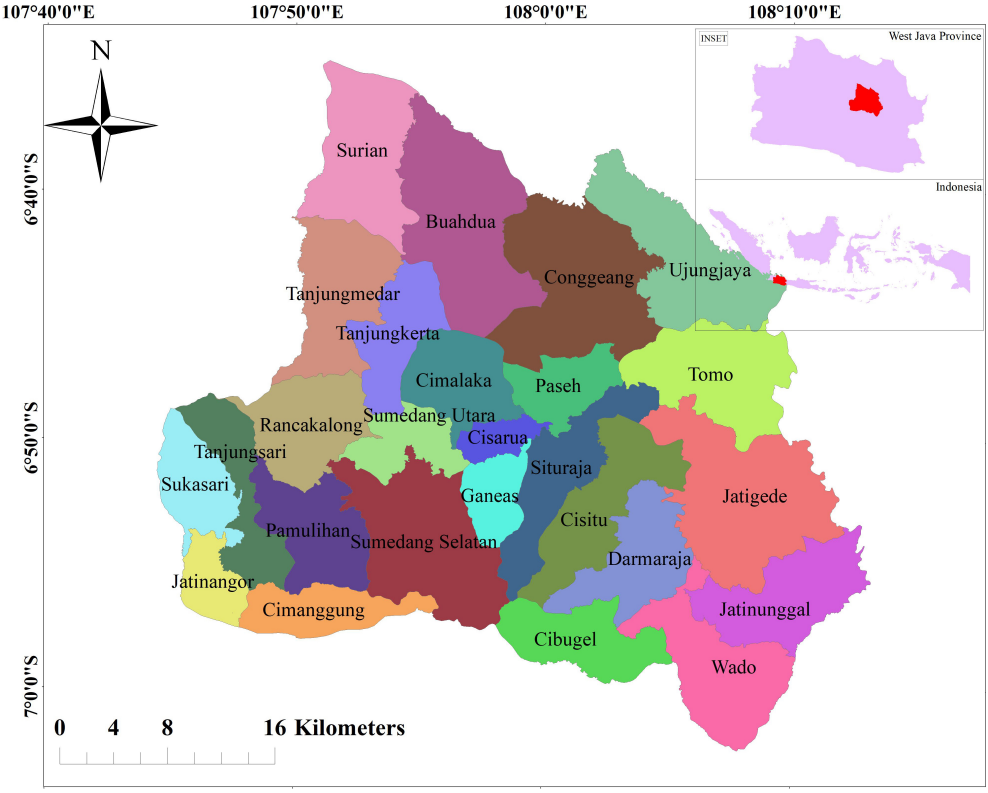

Fig. 1. Sumedang area

Landslide Inventory

The accuracy of the detailed description of an area requires an inventory of landslide events so that it can properly provide information on the characteristics of landslides [12, 35, 36]. In this study, the landslide inventory was obtained from 104 locations of all landslide events from 2000 to 2021, which were inventoried from various references such as field notes from the National Institute of Aeronautics and Space (LAPAN), disaster reports of the disaster-management authority of the National Agency for Disaster Countermeasure (BNPB) and the Center of Volcanology and Geological Hazard Mitigation (PVMBG).

Moreover, the study area has frequently experienced landslides in recent years. In 2021, there were 22 landslide events; most of these occurrences were triggered by high rainfall and human activities. According to Figure 2, it can be seen that most of the landslide disasters were in the middle to the south of the study area.

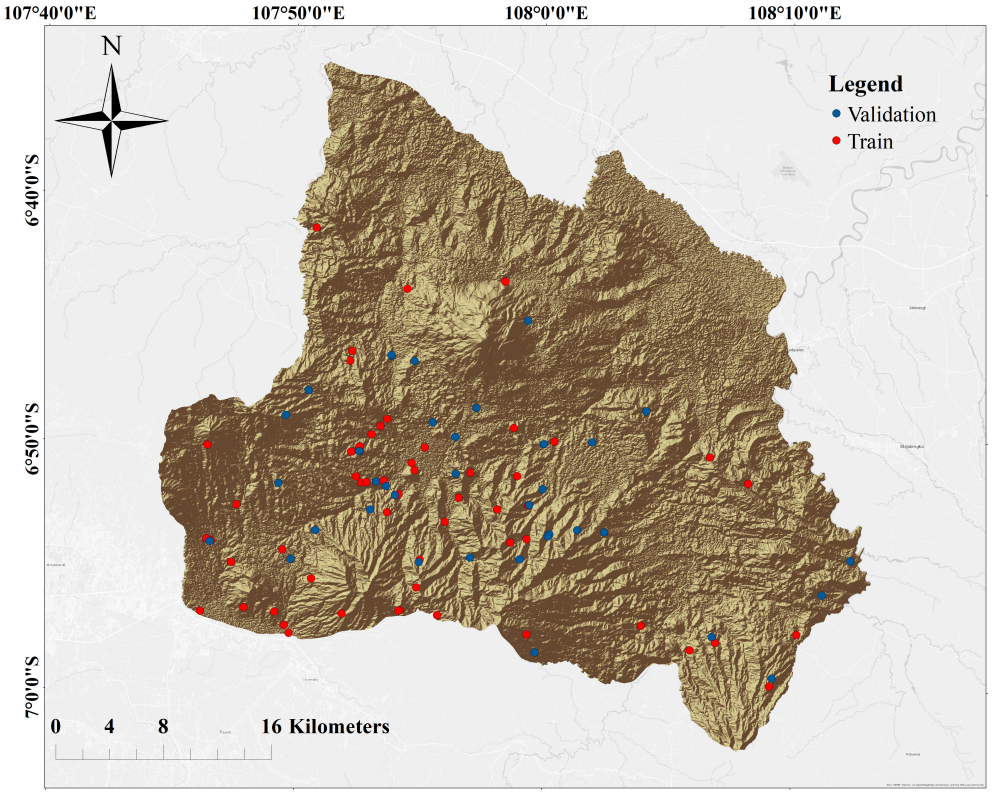

Fig. 2. Landslide inventory on study area

Landslide Factor

Landslides depend not only on the geological conditions of the area but also on other factors such as climate, hydrology, land cover, and vibrations. The landslide-susceptibility model is essential for selecting landslide-triggering and landslide-driving factors. Based on the characteristics and history of the disasters in the research area, this study chose 12 factors for building a landslide-susceptibility model in the research area. Among these, there are topographical and geological factors, the driving factors for landslide disasters and climatic factors, and hydrology and human activities as the triggering factors that were presented in Figure 3 (on the interleaf). Finally, the following information in Table 1 includes the landslide-disaster factors (along with the sources) that were used in this study.

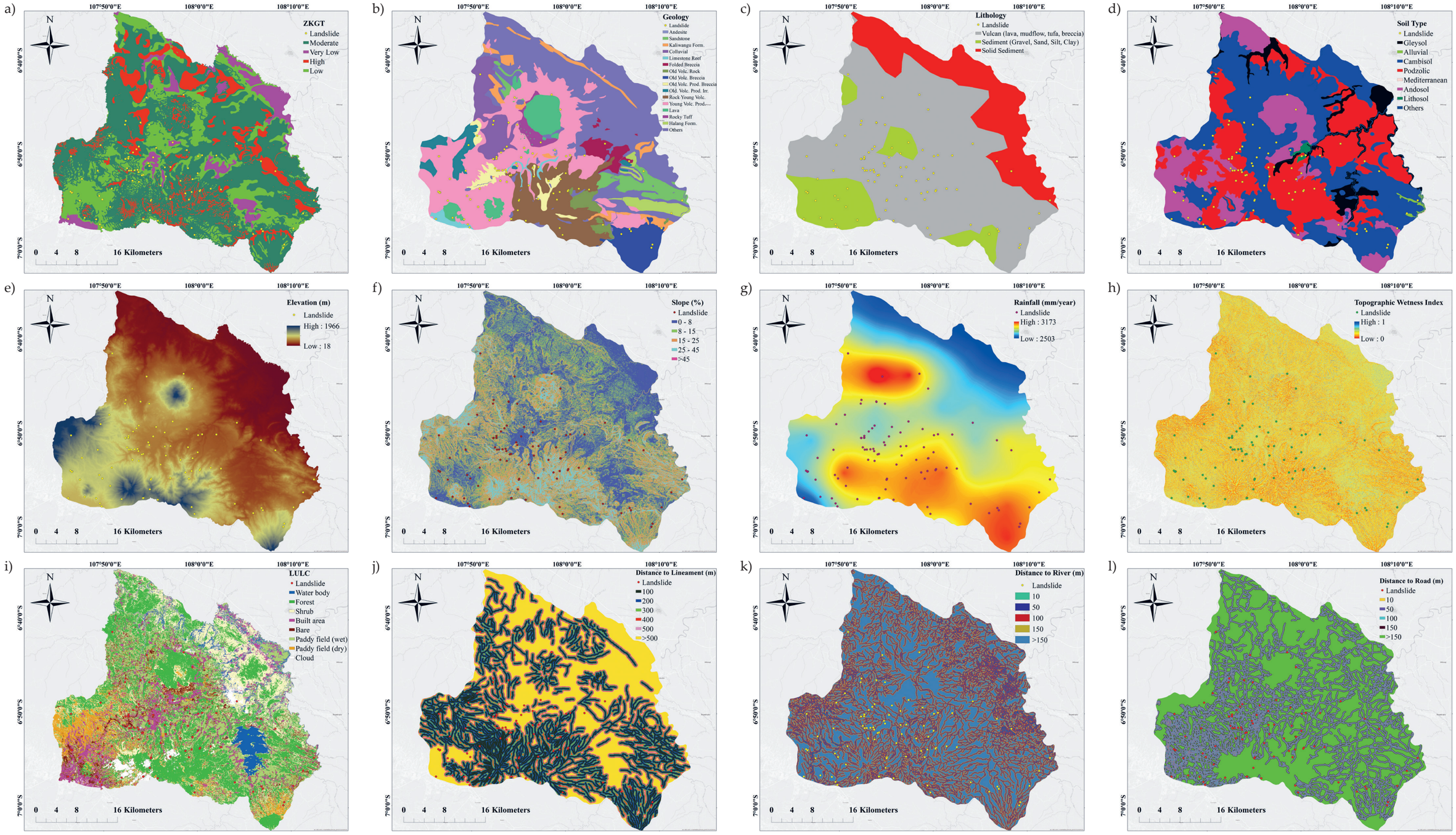

Fig. 3. Data set map: a) ZKGT; b) geology; c) lithology; d) soil type; e) elevation; f) slope; g) rainfall; h) TWI; i) LULC; j) distance to lineament; k) distance to river; l) distance to road

Table 1. Landslide-susceptibility parameters

No.	Factor	Format	Resolution [m]	Source
1	ZKGT	polygon	–	Indonesian Ministry of Energy and Mineral Resources
2	Geology	polygon	–	Indonesian Ministry of Energy and Mineral Resources
3	Lithology	polygon	–	[37–39]
4	Soil type	polygon	–	Indonesian Ministry of Agriculture
5	Elevation and slope	raster	8	Digital Elevation Model (DEM)
6	Rainfall	raster	5500	Climate Hazards Group InfraRed Precipitation with Station (CHIRPS)
7	TWI	raster	8	DEM
8	LULC	raster	30	Landsat 8
9	Distance to lineament	line vector	–	DEM
10	Distance to road & distance to river	line vector	–	Base map from Indonesia Geospatial Portal

2.2. Methodology

The best method is hoped to be more efficient and faster in mapping landslides throughout Indonesia. The stages that have been carried out in this study can be seen in Figure 4, including the data preparation, landslide-prone modeling using FR and RF, validation, and comparisons involving ROC curves and AUC values.

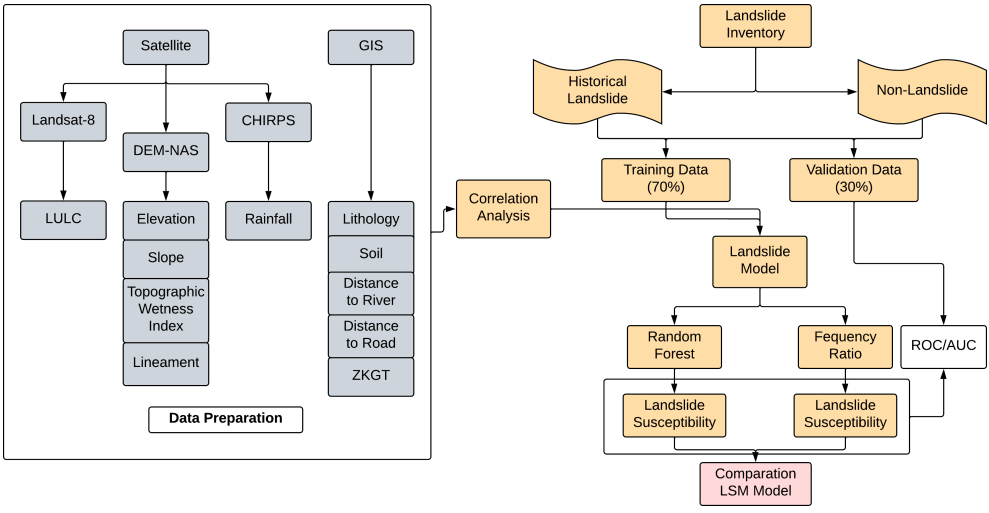

Fig. 4. Methodological framework of study

Preparation Data

When predicting and analyzing objects that can approach reality in the field, the primary and essential thing is to add variables and data accuracy so that users can use it. The data that was used in this study consisted of all of the parameters in Table 1. All the data were then converted into raster format and its spatial resolution was then made uniform into a size of 8.3 meters. In both FR and RF, 70% of the sample was used for training, and 30% was used for testing. The random forest function in the Python scientific kit package was used to develop the RF model with these training data sets.

Frequency Ratio (FR)

FR is a technique that is derived from a ratio approach; our focus was where landslides have occurred throughout the study area. FR is the ratio of the probability that landslides will not occur, as FR can clearly describe the difference in each score among the classes of landslides and landslide events [40, 41]. FR was calculated as the ratio of landslide occurrences in the class to the total area of the class. If the FR index was around 1, it had a neutral effect on the potential for landslides. In addition, the prediction rate (PR) was carried out to prove the relativity of each spatial factor with the available training samples. The landslide-susceptibility map (LSM) was obtained by adding up the multiplication results of all of the factors (which were reclassified) with their respective PR values. The results were divided into five classes: very low, low, moderate, high, and very high landslide susceptibility. Figure 5 shows the landslide-susceptibility framework in the study area.

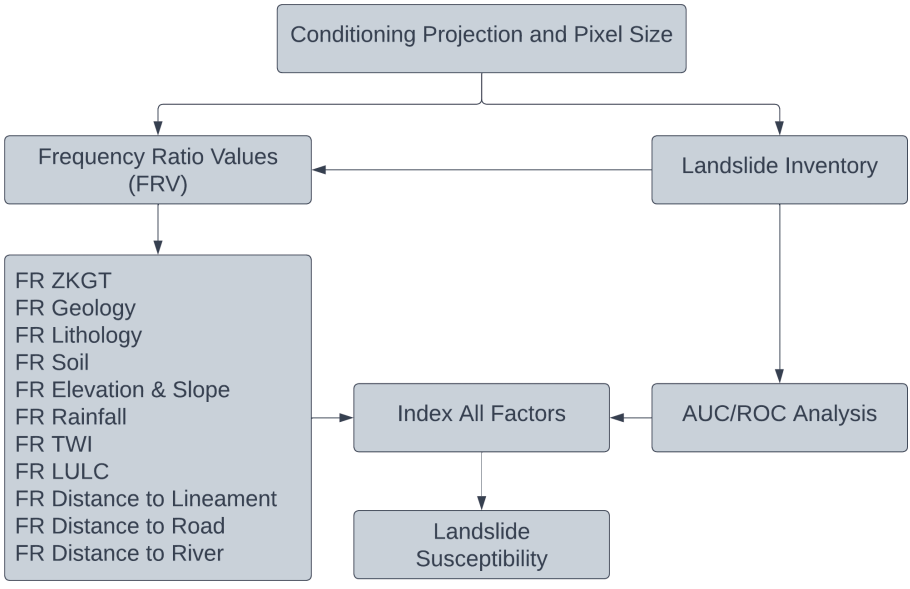

Fig. 5. Schematic FR for landslide-susceptibility map

Random Forest (RF)

RF is a combination of the classification of several data points that consists of a decision tree as an authorized capital [42]. The decision tree is a predictive model that uses a binary rule to determine a target variable [33]. The voting method obtains predictive data from independent data sets that have been trained into the decision tree by resampling. The aim is to collect homogeneous data that is hidden in a large volume of data as the probability of landslide occurrence. This study divides the probability results into several classes according to the threshold. The best sample is selected as a random subset of the training data (as variables) in order to measure the level of correlation among the variables and the results using the Gini criteria [42]. The smaller the Gini index, the smaller the probability that the selected sample in the set will be classified incorrectly. On the other hand, the higher the Gini index, the lower the purity of the set.

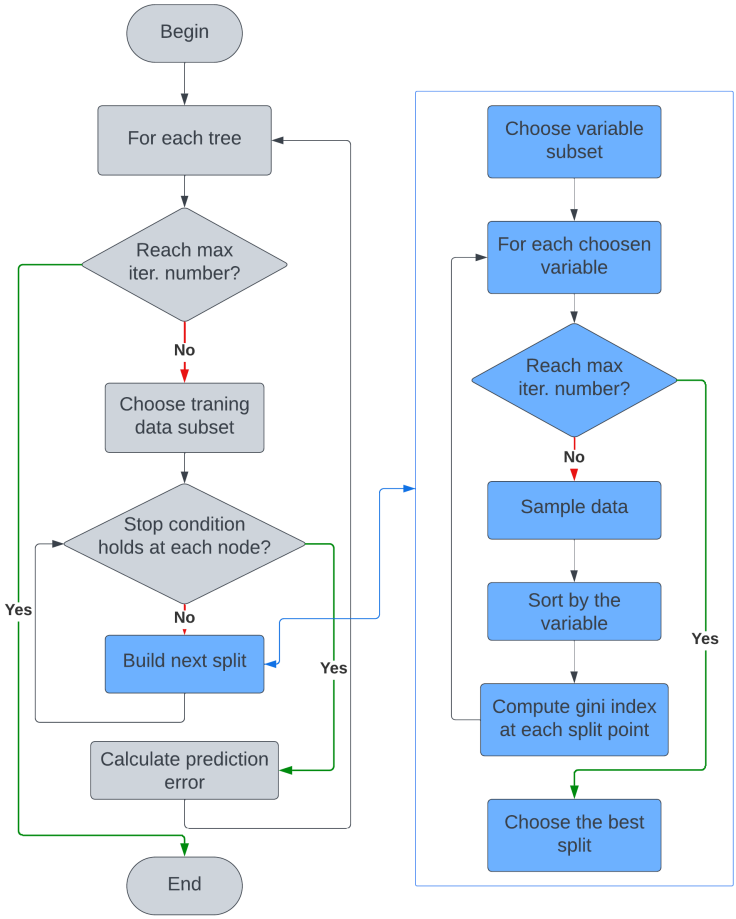

Fig. 6. Schematic RF for landslide-susceptibility map

In RF, it is also possible to determine the critical factor variables for ranking the essential factors in the model using mean decrease impurity (MDI). MDI is based on the total decrease in the node impurity from splitting on the variable and then averaged over all of the trees [43]. Figure 6 shows how the RF algorithm works.

Model Performance Evaluation

Each model was analyzed in order to validate the performance of the landslide-susceptibility mapping by means of the area under the ROC curve (AUC). ROC is a kind of curve that is based on a confusion matrix (which takes sensitivity and specificity as the horizontal and vertical axes, respectively [32]), while AUC represents the accuracy of the model. When the AUC value is close to 1, the prediction accuracy of the model is higher (and vice versa). According to [12], AUC values are categorized as being poor (0.5–0.6), average (0.6–0.7), good (0.7–0.8), very good (0.8–0.9), and excellent (0.9–1.0); however, it can be simplified that the landslide model achieves good performance for landslide-susceptibility assessment if its AUC values are greater than 0.8 [12]. A better model and an AUC value of 1 indicate a perfect model [34]. To build AUC, the following measures are incorporated:

- *sensitivity* = $T_p/(T_p + F_n)$ (percentage of positive cells that were correctly identified),
- *specificity* = $T_n/(T_p + F_n)$ (negative percentage of correct cells identified),

where F_p is a false positive, F_n is a false negative, T_p is truly positive, and T_n is truly negative.

In addition to, the McNemar test is assigned to determine whether both procedures produce the same performance or not. The McNemar test is a well-known statistical technique for determining the statistical significance of discrepancies in classifier performance [44]. The test is a chi-square test for goodness of fit, comparing the distribution of the expected counts under the null hypothesis to the observed counts [45]. It is applied to a 2 × 2 contingency table, with cells containing the numbers of samples that are correctly and wrongly recognized by both techniques as well as the number of samples that are correctly categorized by only one approach [45]. When comparing two binary classification methods, the test comments on whether the two models disagree (or not); it makes no judgment about whether one model is more or less accurate or more prone to errors than the other. The test's null hypothesis (or default assumption) states that the two methods differ by the same amount. If the null hypothesis is rejected, this implies that there is evidence that indicates that the methods differ in various ways.

3. Result

3.1. Application of Frequency Ratio Model

The FR value of each class for each conditioning factor in the study is shown in Table 2.

Table 2. FR for each class of conditioning factors

Class	Number of Pixels in Each Class	Total Pixels Area	Number of Landslide Pixels	Frequency Ratio (FR)	Prediction Rate (PR)
Land Movement Vulnerable Zone (ZKGT)					
Very low	17,241,167	22,612,880	3	0.43	1.65
Low	1,945,776		22	0.96	
Moderate	5,360,522		52	0.89	
High	9932		16	1.17	
Geology					
Andesite	13,664	22,887,214	1	16.42	2.47
Sandstone	851,190		12	3.16	
Halang formation	482,282		1	0.47	
Rock volcanic young	2,530,552		20	1.77	
Limestone reef	239,393		4	3.75	
Old volcanic rock breccia	1,023,356		5	1.10	
Old volcanic rock product breccia	509,850		7	3.08	
Folded breccia	477,759		3	1.41	
Kaliwangu formation	754,499		1	1.29	
Old volcanic rock	374,920		1	0.59	
Colluvial	1,223,349		10	1.83	
Lava	1,078,535		2	0.42	
Young volcanic product	5,767,498		33	1.28	
Old volcanic products are irreversible	584,971		1	0.38	
Rocky tuff	275,507		1	0.81	
Others	6,688,900		0	0.00	
Lithology					
Lava, volcanic mudflow, tuff, breccia	6184	22,887,214	72	1.08	2.06
Solid sediment	4,229,586		3	0.21	
Semi-solid sediment (gravel, sand, silt, clay)	80,565		18	0.86	

Table 2. cont.

Soil Type					
Gleysol	12,056,378	22,887,214	0	0.00	2.46
Alluvial	807,846		0	0.00	
Cambisol	9,077,103		38	0.88	
Podzolic	945,887		40	1.26	
Mediterranean	1,391,467		0	0.00	
Andosol	65,439		13	0.69	
Lithosol	9,742,493		2	5.57	
Others	7,140,751		0	0.00	
Elevation					
Lowland	801,684	22,886,670	0	0.00	2.89
Lowland hills	2,237,488		0	0.00	
Low hills	2,561,232		0	0.00	
Hills	7,209,621		27	0.84	
High hills	10,076,645		66	1.47	
Rainfall [mm/year]					
2380–2685	3,925,465	22,887,837	6	0.34	1.92
2685–2828	6,020,295		25	0.93	
2828–2949	6,691,149		33	1.11	
2949–3173	6,250,928		29	1.04	
TWI					
0	6,796,964	22,601,689	76	0.97	1.73
1	4,563,688		17	0.70	
LULC					
Water body	762,584	22,612,393	0	0.00	1.19
Forest	7,926,169		28	0.78	
Shrubs	4,716,939		10	0.47	
Building area	2,918,509		28	2.13	
Open space area	1,935,478		11	1.26	
Wet agriculture	2,852,108		7	0.54	
Dry agriculture	1,116,982		6	1.19	

Table 2. cont.

Class	Number of Pixels in Each Class	Total Pixels Area	Number of Landslide Pixels	Frequency Ratio (FR)	Prediction Rate (PR)
Distance to Lineament [m]					
100	5,359,551	22,886,661	22	0.92	1.07
200	4,392,428		32	1.64	
300	3,083,147		17	1.24	
400	2,106,579		7	0.75	
500	1,493,026		6	0.90	
>500	6,451,930		18	0.63	
Distance to Road [m]					
10	1,146,622	22,886,661	5	0.98	1.41
50	3,305,992		29	1.97	
100	3,307,232		12	0.81	
150	2,617,581		13	1.11	
>150	12,509,234		34	0.61	
Distance to River [m]					
10	1,472,763	22,886,661	6	0.91	1.00
50	4,516,315		13	0.65	
100	4,799,212		22	1.03	
150	3,684,963		15	0.91	
>150	8,413,408		37	0.99	

According to these FR values, the PR values were derived; the highest PR value was an elevation factor with a value of 2.89, followed by geology and soil type (with values of 2.47 and 2.46), respectively (as shown in Figure 7). By adding up the results of the multiplications of all of the factors with their respective PR values, an LSM was obtained (as shown in Figure 8) where susceptibility was classified into five classes; namely, very low, low, moderate, high, and very high. It can be seen in Table 3 that the share of the very high class was only 0.06%, which was located in the southwest area and center of the study area (Fig. 8). Meanwhile, the percentages of the high, moderate, low, and very low classes were approximately 25, 35, 25, and 15%, respectively. Through the use of the overlay method between the results of the LSM and each factor, it can be noticed that the high to very high LSM classes of rocks in the study area were vulcan with a cambisol soil type with no hydromorphic symptoms. In contrast, the very low to moderate LSM were a sedimentary rock. In addition to, alluvial soil types, andosols, and others have characteristics that are sensitive

to erosion. Furthermore, the FR model on the landslide susceptibility in the study area between the training data that was used in the running model and the vulnerability map showed that the level of prediction had an AUC value of 0.88 (Fig. 9).

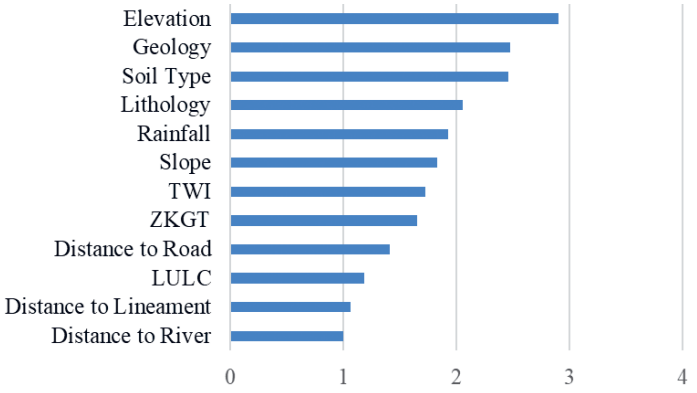

Fig. 7. PR value for each factor

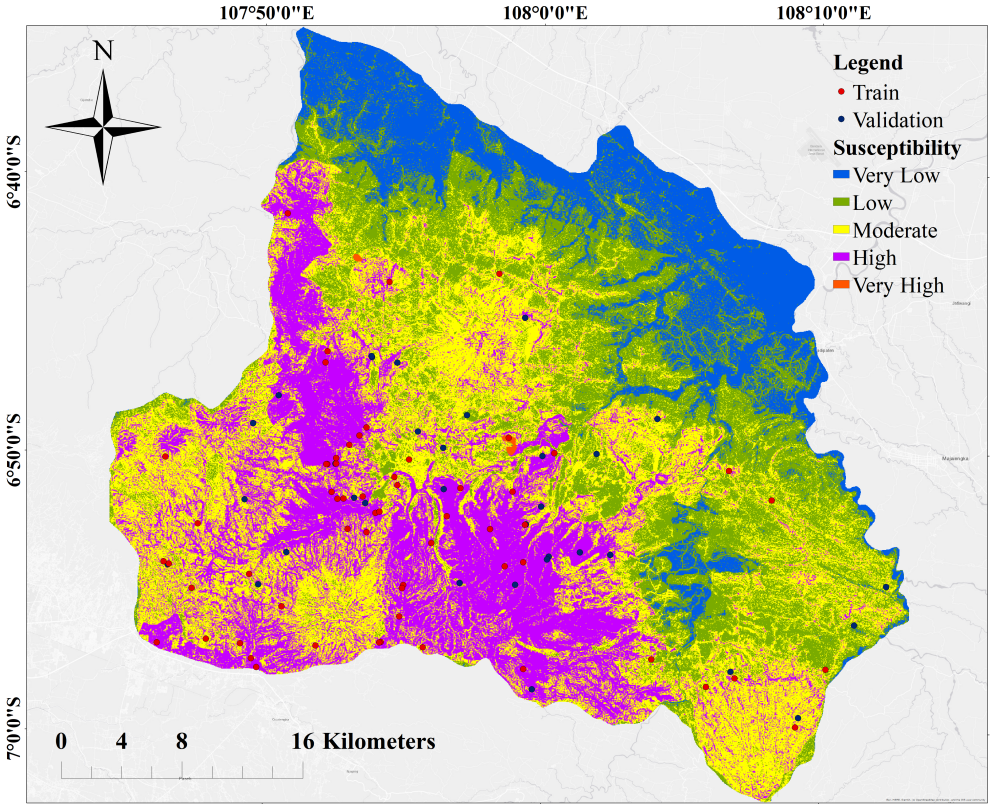

Fig. 8. FR model landslide-susceptibility map

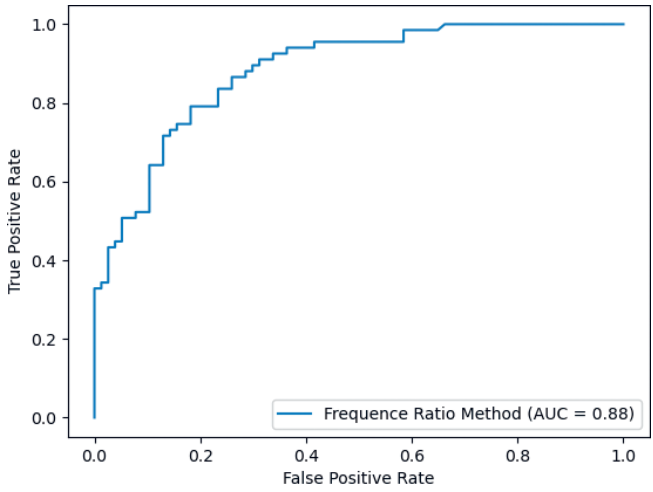

Fig. 9. Area under curve (AUC) of FR model

Table 3. Landslide area

Classes	Area [ha]
Very Low	23.339
Low	39.339
Moderate	55.235
High	39.641
Very High	0.095

3.2. Application of Random Forest Model

After the processes that were explained in Figure 6 were finished, the importance value for each factor (Fig. 10) and the probability map were carried out by means of MDI and RF voting method, respectively. Moreover, the produced probability map was then classified into five classes susceptibility through the use of Jenks natural breaks classification method to create the LSM (Fig. 11). Finally, using the trained RF model and the test data, the AUC value was then obtained (Fig. 12).

Based on Figure 10, it is known that geology is the most-critical factor in the occurrence of landslides in the study area, followed by LULC and elevation (with importance values of 0.17, 0.11, and 0.11, respectively). Besides this, ZKGT is the least-critical factor, followed by lithology and slope (with significant values of 0.03, 0.04, and 0.06, respectively). Overall, the order of the conditioning factors from the most significant to the least were geology, LULC, elevation, distance to lineament, TWI, distance to road, distance to river, soil type, rainfall, slope, lithology, and ZKGT. The LSM in Figure 11 shows a susceptibility map that is classified into five classes namely very low, low, moderate, high and very high landslide susceptibility. Based

on Table 4, the RF model shows that just over 25% of the study area was in the very low class, just below 20% was in the low class, a little more than 23% was in the moderate class, roughly 20% was in the high class, and approximately 11% was in the very high class. As shown in Figure 12, this model had an AUC value of 0.81.

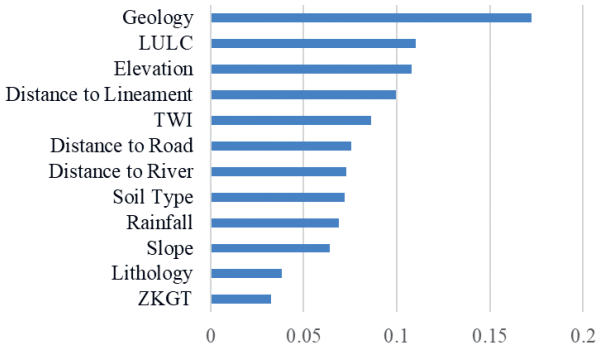

Fig. 10. Importance value for each factor

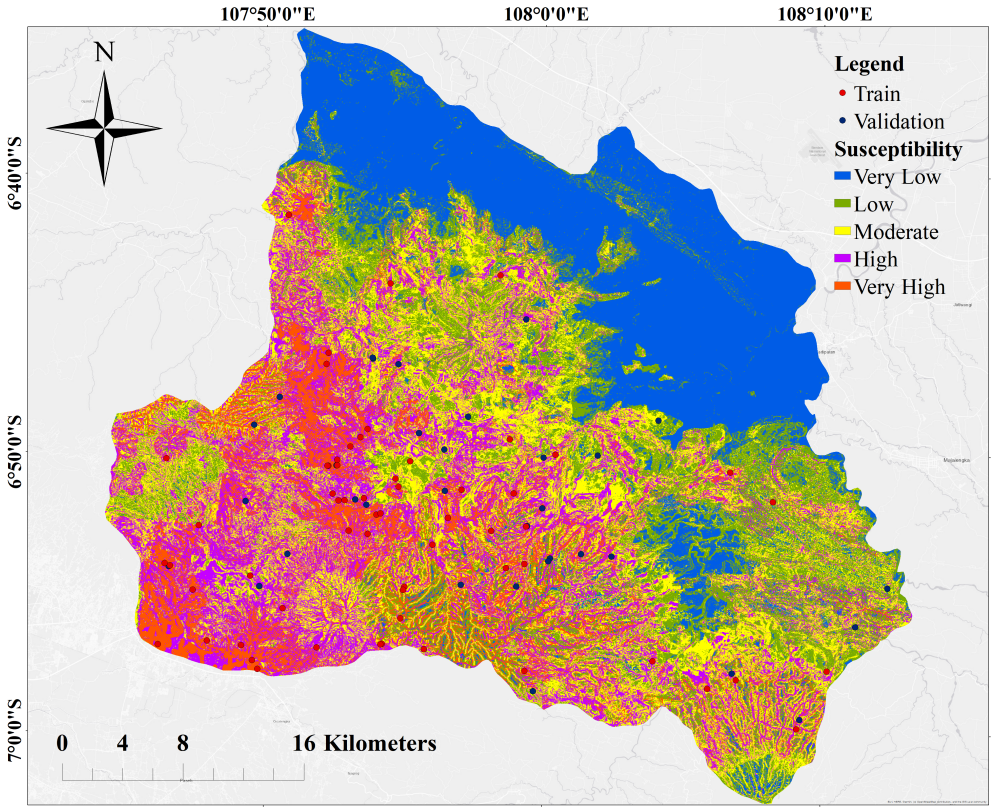

Fig. 11. RF model landslide-susceptibility map

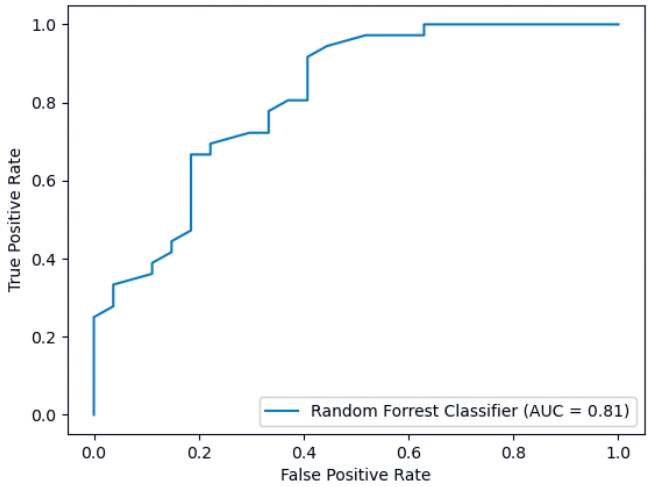

Fig. 12. Area under curve (AUC) of RF model

Table 4. Landslide-potential area

Classes	Area [ha]
Very Low	40.301
Low	31.218
Moderate	36.801
High	31.848
Very High	17.497

4. Comparison and Discussion

4.1. Discussion

Frequency Ratio Model

The FR for each factor class was calculated, followed by its PR. The higher the FR value, the stronger the association with landslides and conditioning factors. For instance, Table 2 shows that the spatial correlation between ZKGT (a parameter relating to ground motion) and landslides which states that the highest FR value for the high class is an area with a generally high rate of ground motion. It shows that a high rate ground motion class in ZKGT parameter has strong correlation to landslide events. Interestingly, in a low class with a low level of ground motion where ground movements rarely occur (unless it is disturbed on a slope), it has the second-highest FR value after the high class; this requires a more detailed study. The geological formation in the study area was the Lembang–Kendeng–Baribis formation, where the

morphological characteristics were low hills with wavy morphology, with heights that ranged from 50 to 200 meters. The east-west line reflected the presence of folds and upward faults. Geologically, the highest FR value was volcanic with andesite rock – closely correlated with the lithosol soil type (this has shallow parent characteristics and is often seen as solid rock on the surface). Geologically, this was a volcanic area, so, the study area was unsurprisingly dominated by volcanic rocks, with the lithosol soil type having the highest FR value. Topographically, the FR value was very well correlated; the higher the hills, the higher the FR value (and vice versa). Elevation will be related to the type of vegetation and rainfall events; the higher the area, the more frequent the rainfall will be, as the cooler weather and atmosphere cannot hold as much condensation (which leads to more-frequent rainfall). The slope will also affect soil strength and, in some instances, lead to landslide potential [5], with higher slope values (especially those with slopes that are greater than 45%) causing rainwater to potentially erode the soil (thus, resulting in potential landslides). In accordance with [4] and [46], the degree of the slope of an area shows a strong relationship with landslides. Some types of soil that easily absorb water and with a certain thickness also have the potential to cause landslides. The main trigger factor for landslides in the study area was rainfall. The average annual rainfall for each grid had various correlations and was dominated by the value of FR 1 in each rainfall class. The land-cover factor positively correlated with building area, open space area, and dry agriculture. The higher the pavement, the higher the runoff when it rains; so, the FR value will be higher, and less vegetation will trigger landslides due to water saturation through heavy rain. While the vegetated land cover had an index of less than 1, this demonstrated the real impact of human activities on landslides. The closer to a lineament, the more the FR value was positively correlated.

Furthermore, the FR model on landslide susceptibility in the study area between the training data that was used in the running model and the vulnerability map showed the level of prediction with an AUC of 0.88 (Fig. 9). According to [12], the obtained value categorized this landslide model as good. However, this result still needs to be improved. With this kind of performance, LSM mapping can detect landslide locations early.

Random Forest Model

Based on Figure 10, it is known that geology was the most critical factor in the occurrence of landslides in the study area, and the least important was ZKGT. The susceptibility classes were moderate to high based on the overlay between the RF model and the geological map. The geological formations were young volcanic rocks with volcanic deposits, coarse, and breccia deposits. The LULC for the very high class was dominated by land cover in the form of built-up and open lands, while the land cover was in the forms of built-up land and forests in the high class. The moderate class was land cover in the forms of forests and shrubs, while the land cover was in the form of shrubs for the very low class.

Next, look at the LSM; the south-to-southwest part of the study area was in the high to very high class, the north and east parts were in the very low class, and the southeast was in the low to moderate class. If viewed from the AUC value (Fig. 12), the obtained value categorized this landslide model as good (but it still needs improvement according to [12]).

4.2. Comparison

To compare whether both models had similar or different performances, a McNemar test was conducted. The hypotheses of the test were as follows:

- H0: FR and RF had the same performances in determining the level of landslide vulnerability in Sumedang.
- H1: FR and RF had different performances in determining the level of landslide vulnerability in Sumedang.

There were 63 test data points applied to both LSM-based FR and RF models to perform the McNemar test. The contingency table for the FR and RF model is shown in Table 5. Using the *mcnemar statsmodels* function on Python, the p-value was derived from the contingency table. The derived *p*-value was 0.26. The null hypothesis was not rejected with a significance level alpha = 0.05, because the *p*-value was more significant than alpha. Therefore, it can be concluded that the FR and RF models performed similarly in determining the landslide-vulnerability levels in Sumedang.

Table 5. Contingency of FR and RF for McNemar test

	RF correct	RF incorrect
FR correct	33	7
FR incorrect	13	10

FR and RF are parts of several existing methods for analyzing landslides. The two methods that were used are both excellent for analyzing landslides in various areas. Whether or not these two methods are superior depends on which perspective we look at; for this case study, both FR and RF were superior because they had AUC performances above 0.80. However, these two methods had their advantages and disadvantages in producing LSMs in the current study area. The results of the LSM using the moderate-class FR method were slightly higher than those of the high class, while the RF model resulted from the very-low-class LSM more than it did from the other classes. Nevertheless, these two methods provide good performance with limited input data – especially if the field data that is used for validation has a more significant amount than what is currently used.

This difference in the results was not a debate about the future because, with the higher awareness, the very low class also needs attention so that landslides do not cause many losses. Suppose only a very low class can control/manage landslides

properly; in this case, the other classes will be better for minimizing landslides in this area in terms of casualties, infrastructure, and the environment. The results of these two methods need to be disseminated properly, as well as the field verification of each class of the LSM. In the future, these two methods must also be tested and analyzed for sub-district scale areas at a scale of 1:5000. Input data with the same scale supports this. With a more detailed scale and better input data, it is hoped that it will clear any doubts about the two methods that are currently being used at this location.

5. Conclusion

The FR and RF methods performed well and had the same performances (according to McNemar test) in analyzing LSMs in the study area. FR and RF were two methods that were both superior, as they each had an AUC performance above 0.80. The results of the LSMs with the moderate class FR method were slightly higher than those of the high class, while the RF model resulted from the very-low-class LSMs more than it did from the other classes. More-extensive field data will provide better performance for the two methods. Statistical methods and machine learning can be compared and used as references in landslide prevention and development planning.

Author Contributions

Hana Listi Fitriana (35%): concept, methodology, data processing, analysis, and writing section.

Rido Dwi Ismanto (25%): methodology section, data processing, analysis, and writing section.

Jessica Stephanie Tulus (15%): data-processing section.

Johannes Manalu (10%): analysis section.

Atriyon Julzarika (10%): review and analysis section.

Jalu Tejo Nugroho (5%): review and analysis section.

Hana Listi Fitriana, Rido Dwi Ismanto, Atriyon Julzarika, and Jalu Tejo Nugroho were the main contributors to this paper, while Johannes Manalu and Jessica Stephanie Tulus were secondary contributors.

Declaration of Competing Interests

We all state that none of us have any competing interests with this article.

Acknowledgment

The authors would like to thank the National Research and Innovation Agency (BRIN), the Regional Agency for Disaster Management (BPBD) of Sumedang, the Indonesian Ministry of Agriculture, and the Productive Innovative Research of Indonesian Ministry of Finance (RISPRO) for supporting this research. We also thank the Climate Hazard Center at the University of California-Santa Barbara (CHC UCSB) for providing the open CHIRPS rainfall data for this research.

Reference

[1] Fathani T.F., Legono D., Karnawati D.: *A numerical model for the analysis of rapid landslide motion*. Geotechnical and Geological Engineering, vol. 35(5), 2017, pp. 2253–2268. https://doi.org/10.1007/s10706-017-0241-9.

[2] Marelyn Telun D., Tham Fatt N., Mohd Farid A.K., Pereira J.J.: *Landslide susceptibility modeling using a hybrid bivariate statistical and expert consultation approach in Canada Hill, Sarawak, Malaysia*. Frontiers in Earth Science, vol. 9, 2021, pp. 1–15. https://doi.org/10.3389/feart.2021.616225.

[3] Salehpour Jam A., Mosaffaie J., Tabatabaei M.R.: *Raster-based landslide susceptibility mapping using compensatory MADM methods*. Environmental Modelling & Software, vol. 159, 2023, 105567. https://doi.org/10.1016/j.envsoft.2022.105567.

[4] Moazzam M.F.U., Vansarochana A., Boonyanuphap J., Choosumrong S., Rahman G., Djueyep G.P.: *Spatio-statistical comparative approaches for landslide susceptibility modeling: Case of Mae Phun, Uttaradit Province, Thailand*. SN Applied Sciences, vol. 2(3), 2020, 384. https://doi.org/10.1007/s42452-020-2106-8.

[5] Dou J., Yunus A.P., Bui D.T., Merghadi A., Sahana M., Zhu Z., Chen C.-W., Khosravi K., Yang Y., Pham B.T.: *Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan*. Science of The Total Environment, vol. 662, 2019, pp. 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221.

[6] Naryanto H.S.: *Analisis Kejadian Bencana Tanah Longsor Banjarnegara, Provinsi Jawa Tengah the December 12, 2014 Landslide Disaster Analysis in Jemblung Area, Sampang Village, Karangkobar Subdistrict, Banjarnegara District, Central Java*. Alami: Jurnal Teknologi Reduksi Risiko Bencana, vol. 1(1), 2017, pp. 1–10. https://doi.org/10.29122/alami.v1i1.122.

[7] Muriyatmoko D., Utama S.N., Pradhana F.R., Umami J., Rozaqi A.J., Setyaningrum H.: *Landslide prediction model of prone areas in Pulung, Ponorogo East Java*. Procedia Computer Science, vol. 161, 2019, pp. 747–755. https://doi.org/10.1016/j.procs.2019.11.179.

[8] Baral N., Karna A.K., Gautam S.: *Landslide susceptibility assessment using modified frequency ratio model in Kaski District, Nepal*. International Journal of Engineeringand Management Research, vol. 11(1), 2021, pp. 167–177. https://doi.org/10.31033/ijemr.11.1.23.

[9] Pradhan B., Mansor S., Pirasteh S., Buchroithner M.F.: *Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model*. International Journal of Remote Sensing, vol. 32(14), 2011, pp. 4075–4087. https://doi.org/10.1080/01431161.2010.484433.

[10] Vineetha P., Sarun S., Sheela A.M.: *Landslide susceptibility analysis using Frequency Ratio Model in a Tropical Region, South East Asia*. Journal of Geography, Environment and Earth Science International, vol. 22(2), 2019, pp. 1–13. https://doi.org/10.9734/jgeesi/2019/v22i230140.

[11] Fathani T.F., Syah A., Faris F.: *A numerical analysis of landslide movements considering the erosion and deposition along the flow path*. Journal of the Civil Engineering Forum, vol. 5(3), 2019, 187. https://doi.org/10.22146/jcef.43808.

[12] Chen W., Peng J., Hong H., Shahabi H., Pradhan B., Liu J., Zhu A.-X., Pei X., Duan Z.: *Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China*. Science of The Total Environment, vol. 626, 2018, pp. 1121–1135. https://doi.org/10.1016/j.scitotenv.2018.01.124.

[13] Nhu V.-H., Mohammadi A., Shahabi H., Ahmad B.B., Al-Ansari N., Shirzadi A., Clague J.J., Jaafari A., Chen W., Nguyen H.: *Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment*. International Journal of Environmental Research and Public Health, vol. 17(14), 2020, 4933. https://doi.org/10.3390/ijerph17144933.

[14] Wang H., Zhang L., Yin K., Luo H., Li J.: *Landslide identification using machine learning*. Geoscience Frontiers, vol. 12(1), 2021, pp. 351–364. https://doi.org/10.1016/j.gsf.2020.02.012.

[15] Wang Y., Sun D., Wen H., Zhang H., Zhang F.: *Comparison of random forest model and frequency ratio model for Landslide Susceptibility Mapping (LSM) in Yunyang County (Chongqing, China)*. International Journal of Environmental Research and Public Health, vol. 17(12), 2020, 4206. https://doi.org/10.3390/ijerph17124206.

[16] Zhou X., Wen H., Zhang Y., Xu J., Zhang W.: *Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization*. Geoscience Frontiers, vol. 12(5), 2021, 101211. https://doi.org/10.1016/j.gsf.2021.101211.

[17] Azarafza M., Azarafza M., Akgün H., Atkinson P.M., Derakhshani R.: *Deep learning-based landslide susceptibility mapping*. Scientific Reports, vol. 11(1), 2021, 24112. https://doi.org/10.1038/s41598-021-03585-1.

[18] Hong Y., Adler R., Huffman G.: *Use of satellite remote sensing data in the mapping of global landslide susceptibility*. Natural Hazards, vol. 43(2), 2007, pp. 245–256. https://doi.org/10.1007/s11069-006-9104-z.

[19] Xing Y., Yue J., Guo Z., Chen Y., Hu J., Travé A.: *Large-scale landslide susceptibility mapping using an integrated machine learning model: A case study in the Lvliang Mountains of China*. Frontiers in Earth Science, vol. 9, 2021, pp. 1–15. https://doi.org/10.3389/feart.2021.722491.

[20] Shirzadi A., Soliamani K., Habibnejhad M., Kavian A., Chapi K., Shahabi H., Chen W., Khosravi K., Thai Pham B., Pradhan B., Ahmad A., Ahmad B.B., Bui D.T.: *Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping*. Sensors, vol. 18(11), 2018, 3777. https://doi.org/10.3390/s18113777.

[21] Akinci H., Kilicoglu C., Dogan S.: *Random forest-based landslide susceptibility mapping in coastal regions of Artvin, Turkey*. ISPRS International Journal of Geo-Information, vol. 9(9), 2020, 553. https://doi.org/10.3390/ijgi9090553.

[22] Abdo H.G.: *Assessment of landslide susceptibility zonation using frequency ratio and statistical index: a case study of Al-Fawar basin, Tartous, Syria*. International Journal of Environmental Science and Technology, vol. 19(4), 2022, pp. 2599–2618. https://doi.org/10.1007/s13762-021-03322-1.

[23] Sørensen R., Zinko U., Seibert J.: *On the calculation of the topographic wetness index: Evaluation of different methods based on field observations*. Hydrology and Earth System Sciences, vol. 10(1), 2006, pp. 101–112. https://doi.org/10.5194/hess-10-101-2006.

[24] Efendi D., Hidayah E., Hasanuddin A.: *Mapping of landslide susceptible zones by using frequency ratios at Bluncong subwatershed, Bondowoso Regency*. U KaRsT, vol. 5(1), 2021, pp. 126–141. https://doi.org/10.30737/ukarst.v5i1.1455.

[25] Hong L., Ouyang M., Peeta S., He X., Yan Y.: *Vulnerability assessment and mitigation for the Chinese railway system under floods*. Reliability Engineering & System Safety, vol. 137, 2015, pp. 58–68. https://doi.org/10.1016/j.ress.2014.12.013.

[26] Javier D.N., Kumar L.: *Frequency ratio landslide susceptibility estimation in a tropical mountain region*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-3/W8, 2019, pp. 173–179. https://doi.org/10.5194/isprs-archives-XLII-3-W8-173-2019.

[27] Kose D.D., Turk T.: *GIS-based fully automatic landslide susceptibility analysis by weight-of-evidence and frequency ratio methods*. Physical Geography, vol. 40(5), 2019, pp. 481–501. https://doi.org/10.1080/02723646.2018.1559583.

[28] Nicu I.C.: *Frequency ratio and GIS-based evaluation of landslide susceptibility applied to cultural heritage assessment*. Journal of Cultural Heritage, vol. 28, 2017, pp. 172–176. https://doi.org/10.1016/j.culher.2017.06.002.

[29] Oh H.-J., Lee S., Hong S.-M.: *Landslide susceptibility assessment using frequency ratio technique with iterative random sampling*. Journal of Sensors, vol. 2017, 2017, 3730913. https://doi.org/10.1155/2017/3730913.

[30] Darminto M.R., Widodo A., Alfatinah A., Chu H.J.: *High-resolution landslide susceptibility map generation using machine learning (case study in Pacitan, Indonesia)*. International Journal on Advanced Science, Engineering and Information Technology, vol. 11(1), 2021, pp. 369–379. https://doi.org/10.18517/ijaseit.11.1.11679.

[31] Park S., Kim J.: *Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance*. Applied Sciences, vol. 9(5), 2019, 942. https://doi.org/10.3390/app9050942.

[32] Sun D., Wen H., Wang D., Xu J.: *A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm*. Geomorphology, vol. 362, 2020, 107201. https://doi.org/10.1016/j.geomorph.2020.107201.

[33] Taalab K., Cheng T., Zhang Y.: *Mapping landslide susceptibility and types using Random Forest*. Big Earth Data, vol. 2(2), 2018, pp. 159–178. https://doi.org/10.1080/20964471.2018.1472392.

[34] Youssef A.M., Pourghasemi H.R., Pourtaghi Z.S., Al-Katheeri M.M.: *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(5), 2016, pp. 839–856. https://doi.org/10.1007/s10346-015-0614-1.

[35] Guzzetti F., Manunta M., Ardizzone F., Pepe A., Cardinali M., Zeni G., Reichenbach P., Lanari R.: *Analysis of ground deformation detected using the SBAS-DInSAR technique in Umbria, Central Italy*. Pure and Applied Geophysics, vol. 166(8–9), 2009, pp. 1425–1459. https://doi.org/10.1007/s00024-009-0491-4.

[36] van Westen C.J.: *The modelling of landslide hazards using GIS*. Surveys in Geophysics, vol. 21(2–3), 2000, pp. 241–255. https://doi.org/10.1023/A:1006794127521.

[37] van Bemmelen R.W.: *General geology of Indonesia and adjacent archipelagoes: The East Indies, inclusive of the British part of Borneo, the Malay Peninsula, the Philippine Islands, Eastern New Guinea, Christmas Island, and the Andaman- and Nicobar Islands*. 1949. https://api.semanticscholar.org/CorpusID:177002077.

[38] Djuri: *Peta geologi lembar Arjawinangun, Jawa.* Pusat Penelitian dan Pengembangan Geologi, Bandung 1973.

[39] Djuri: *Peta geologi lembar Arjawinangun, Jawa.* Pusat Penelitian dan Pengembangan Geologi, Bandung 1995.

[40] Pradhan B., Lee S.: *Landslide risk analysis using artificial neural network model focussing on different training sites*. International Journal of the Physical Sciences, vol. 4(1), 2009, pp. 001–015.

[41] Rasyid A.R., Bhandary N.P., Yatabe R.: *Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia*. Geoenvironmental Disasters, vol. 3(1), 2016, 19. https://doi.org/10.1186/s40677-016-0053-x.

[42] Breiman L.: *Random forests*. Machine Learning, vol. 45(1), 2001, pp. 5–32. https://doi.org/10.1023/A:1010933404324.

[43] Biau G., Scornet E.: *A random forest guided tour*. TEST, vol. 25(2), 2016, pp. 197–227. https://doi.org/10.1007/s11749-016-0481-7.

[44] Dietterich T.G.: *Approximate statistical tests for comparing supervised classification learning algorithms*. Neural Computation, vol. 10(7), 1998, pp. 1895–1923. https://doi.org/10.1162/089976698300017197.

[45] Kavzoglu T.: *Object-oriented random forest for high resolution land cover mapping using Quickbird-2 imagery*. [in:] Samui P., Sekhar S., Balas V.E. (eds.), *Handbook of Neural Computation*, Academic Press, Cambridge, MA 2017, pp. 607–619. https://doi.org/10.1016/B978-0-12-811318-9.00033-8.

[46] Nakileza B.R., Nedala S.: *Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda*. Geoenvironmental Disasters, vol. 7(1), 2020, 27. https://doi.org/10.1186/s40677-020-00160-0.