International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 2
ISSN: 2088-5334

Utilization of Frequency Ratio and Logistic Regression Model for Landslide Susceptibility Mapping in Bogor Area

Raditya Panji Umbara [a], Dian Nuraini Melati [a], Astisiasari [a], Wisyanto [a], Syakira Trisnafiah [a], Trinugroho [a], Yukni Arifianti [a], Firman Prawiradisastra [a], Taufik Iqbal Ramdhani [b], Samsul Arifin [c,*], Maria Susan Anggreainy [d]

[a] *Research Center for Geological Disaster, National Research and Innovation Agency (BRIN), Bandung, 40135, Indonesia*
[b] *Research Center for Artificial Intelligence and Cyber Security, National Research and Innovation Agency (BRIN), Bandung, Indonesia*
[c] *Data Science, Faculty of Engineering and Design, Institut Teknologi Sains Bandung, West Java,17530 Indonesia*
[d] *Computer Science Department, BINUS Graduate Program Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia*
Corresponding author: [*]samsul.arifin@itsb.ac.id

***Abstract*— Java Island holds the highest record of landslide events in Indonesia. In 2021, the Bogor area, consisting of the city and regency of Bogor, recorded the highest number of landslides. These events further impact the fatalities, damage, and loss to society. Landslide mitigation should be considered to reduce the risk caused by landslide hazards. In this regard, a landslide susceptibility analysis is one of the fundamental steps in mitigation measures that can support policymakers in response to landslide disaster risk reduction. The location of landslide possibilities can be identified by mapping landslide susceptibility. Therefore, this study aims to produce a landslide susceptibility map (LSM) using a statistical frequency ratio method and logistic regression. The number of landslide inventories used in the model is about 822 events. To apply the model, the present study evaluates 13 influencing factors consisting of elevation, slope angle, slope aspect, slope curvature, Topographic Wetness Index (TWI), distance to river, lithological, distance to fault, soil type, annual rainfall, Normalized Difference Vegetation Index (NDVI), land use land cover (LULC), and road distance. The model performance is further evaluated using Area Under the ROC Curve (AUC). The frequency ratio (FR) and logistic regression (LR) models produce satisfactory results and have high predictions of future landslide occurrences with a score of 0.8317 and 0.8817, respectively.**

***Keywords*— Frequency ratio; logistic regression; GIS; landslide susceptibility mapping; Bogor; Indonesia.**

Manuscript received 31 Jul. 2023; revised 9 Jan. 2024; accepted 12 Mar. 2024. Date of publication 30 Apr. 2024.

I. INTRODUCTION

The geological and geographical conditions of Indonesia's territory make it fertile. Still, it is also accompanied by various kinds of natural disaster threats, such as earthquakes, tsunamis, volcanic eruptions, landslides, floods, forest fires, droughts, and hurricanes. Based on historical data on disaster events in Indonesia, since 2005, there has been an increasing trend of disaster events. In 2005, there were 612 disasters, and it continues to increase to 4,977 disasters in 2020 [1]. The most natural disasters in 2020 were floods, tornadoes, and landslides. Among the three disasters, the landslide disaster was the disaster that claimed the most lives. Landslide is the most threatening geological disaster and has an impact on socio-economic growth in many countries [2]–[5].

Several landslides occurred in various regions in Indonesia, namely: (1) Bandung, West Java, on 23 February 2010, which claimed 33 deaths and 11 missings; (2) Banjarnegara, Central Java, on 12 December 2014, with 20 dead and 88 missing, (3) Agam, West Sumatra on 27 January 2013 which claimed 20 deaths, (4) Buru Island, Maluku on 23 July 2010 with 18 deaths, (5) West Bandung, West Java on March 25, 2013, with 14 deaths and 3 missing, (6) Cimanggung, Sumedang Regency, on January 9, 2021 which claimed 40 lives and what recently happened was a landslide which was affected by Garut Earthquake November 21, 2022 and claimed 336 lives. Landslides in Indonesia, ±75 percent occurred on the island of Java [6]. At the provincial level, Central Java and West Java are the 2 highest compared to other regions [7]. The Bogor area has the highest landslide occurrence in West Java. There were 13 fatalities and missing in 2022 [8].

Landslide disaster mitigation can be carried out structurally (physical construction) or non-structurally (spatial planning, law enforcement, disaster education, mapping, and others). Provision of landslide vulnerability maps is also included in a series of landslide disaster mitigation efforts. This landslide vulnerability map can be used as an early warning tool for the community so that vulnerable points can be used for attention, increase awareness, and reduce potential risks. In addition, it can also be used as a basis for preparing a landslide mitigation spatial plan.

Landslide susceptibility can be mapped in various ways, including statistical methods. One of these methods is a statistical analysis of landslide occurrence data using a probabilistic frequency-ratio model. Frequency ratio analysis to obtain predictive maps of landslide susceptibility has been carried out by many researchers, such as [3], [9]–[15]. In this study, the method for preparing landslide susceptibility maps for the Bogor area was carried out by statistical analysis of frequency ratios. The ratio-frequency model is a simple way, and the input, calculations, and output process are easy to understand [16]. Several studies show that making a landslide susceptibility map based on frequency ratio analysis is better than other analytical methods that are compared in their respective studies, such as research conducted by [16]–[19]. Even so, some argue that combined statistical analysis will obtain better results, as Liu et al. [14] did. who uses a combination analysis of frequency ratio, logistic regression, and index of entropy—another research by Hafsa et al. [20] also proved that a combination of frequency ratio and logistic regression could improve the accuracy of the analysis. Logistic regression involves more complex calculations than frequency ratios. However, logistic regression produces better accuracy than frequency ratio [21], [22]. A reason for selecting Bogor as the target research area is that the area has the highest frequency of landslide events in West Java. A better map is needed using the novelty of existing data, while the selection of the frequency ratio analysis method, apart from being simple and the process is easy to understand, also produces better results, superior to some other methods. A reason for selecting Bogor as the target research area is that the area has the highest frequency of landslide events in West Java. A better map is needed using the novelty of existing data. At the same time, the selection of the frequency ratio and logistic regression analysis method, apart from being simple and easy to understand, also produces better results, superior to other methods.

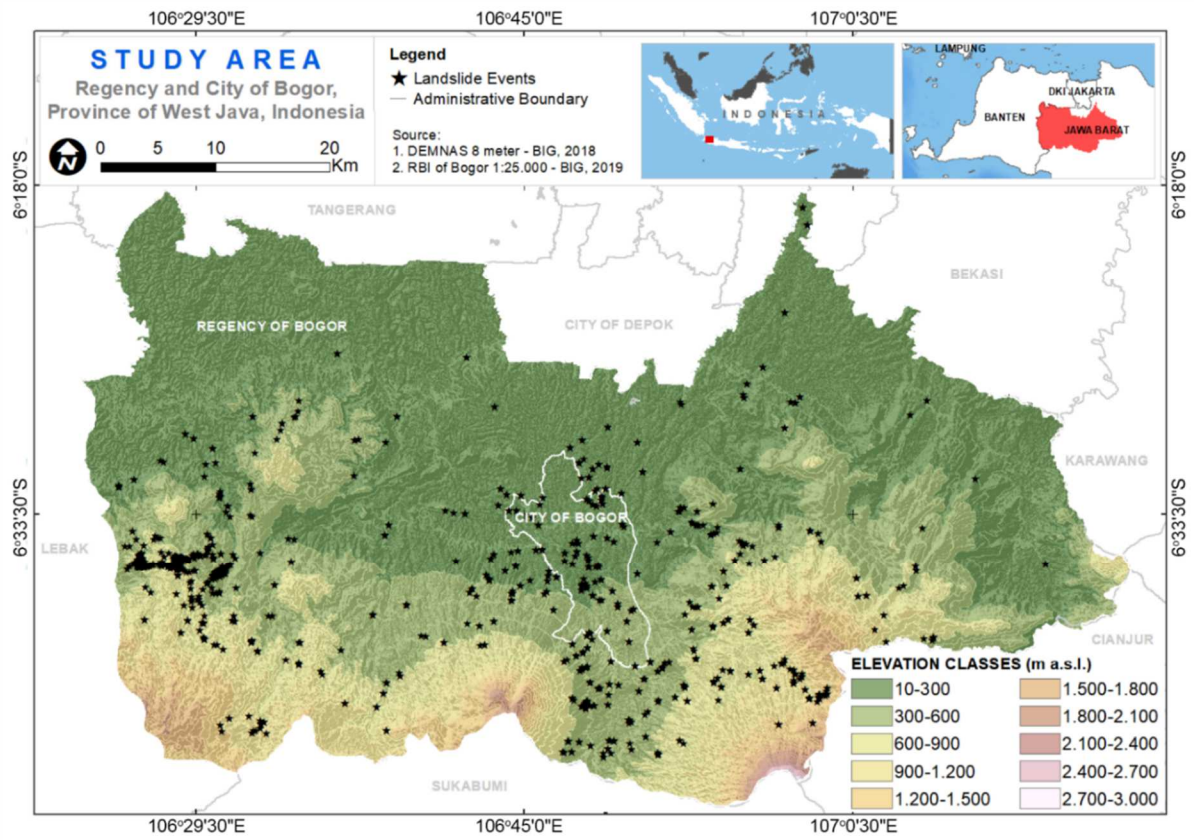


Fig. 1 Study Area

II. Materials and Methods

A. Study Area

Bogor is an administrative area located in West Java, Indonesia. It encompasses both the city of Bogor and Bogor Regency. The population of Bogor, including the city and regency, in 2022 reaches 6,630,350 people, with a population density of 2022 in Bogor City and Bogor Regency for around 9,550 people/km^2 and 1,861 people/km^2, respectively [8]. Bogor City is situated at the foothills of the volcanic Mount Salak and Mount Gede, which are part of the more extensive volcanic range in West Java. In comparison, Bogor Regency encompasses a diverse topography. It includes flat plains,

hills, and mountains. The southern part of Bogor regency is bordered by the mountainous terrain of Mount Salak and Mount Gede to the south. Bogor City receives significant rainfall, with an average annual rainfall of around 3,000-4,000 millimeters. The climate is generally mild, with temperatures ranging from 16 to 28.6 degrees Celsius. The climate in Bogor Regency is similar to that of Bogor City. It has a tropical rainforest climate with high humidity and significant yearly rainfall. The temperature range is between 19.2 and 34.7 degrees Celsius [8].

B. Landslide Inventory

A total of 822 landslide points were mapped in the study area (Fig. 1). These landslide inventories were recorded by the Regional Disaster Management Agency of Bogor for the City and Regency of Bogor. Landslide points were randomly divided into training data (70%) with 575 points and validation data (30%) with 247 points. There are no rules or limits for determining the percentage of training and validation data, generally 80%:20%, 75%:25%, and 70%:30% [23]. However, the consideration in using 70%:30% is that 70% is sufficient to represent data for analysis, while 30% is sufficient to represent data for validation [24].

C. Image Processing

This study used image processing to derive annual rainfall and land use land cover (LULC). To generate rainfall information, CHIRPS data was collected from Google Earth Engine. The acquisition for 2013-2022 was further used to calculate annual rainfall. For LULC, Sentinel-2 image processing was performed. This study used the acquisition of Sentinel-2 images with the least cloud cover in 2021 for further image classification. Object-based image analysis (OBIA) was implemented for image classification. In this case, the Simple Non-Iterative Clustering (SNIC) algorithm provided by Google Earth Engine [25] was applied for image segmentation. Furthermore, image classification was performed using random forest classifiers. This method was found to produce better LULC classification compared to pixel-based image analysis [26].

D. Landslide Influencing Factors

In the present study, thirteen influencing factors are used to generate Landslide Susceptibility Map (LSM), i.e., elevation, slope angle, slope aspect, slope curvature, Topographic Wetness Index (TWI), distance to river, lithological, distance to fault, soil type, annual rainfall, Normalized Difference Vegetation Index (NDVI), LULC, and road distance. All these data are presented as geospatial data with different types of data formats, as shown in Table 1 and Fig. 2. To implement the model of LSM. All the data are then formatted into a raster data type with a pixel size of 25 m.

TABLE I
LANDSLIDE INFLUENCING FACTOR DATA

Topographical Factors	**Description**	**Source**	**Resolution/ Scale**	**Type**
Elevation	Elevation is related to slope stability, precipitation properties, and vegetation coverage.	DEMNAS, Indonesian Geospatial Information Agency (BIG), 2018	0.27 arcsecond	Raster
Slope angle	The slope angle corresponds to soil moisture, surface runoff, and constraint stress.			
Slope aspect	The slope aspect influences the weathering effect, precipitation, and vegetation development.			
Slope curvature	Slope shape is associates with water drainage or material diversion based on its concave or convex shape.			
Topographic Wetness Index (TWI)	TWI reflects the accumulation of water flow, indicating land saturation.			
Distance to river	The probability of pore water pressure might rise in the area around the river. It impacts erosion and slope stability.	Indonesian Geospatial Information Agency (BIG), 2019	1:25,000	Vector
Geological				
Lithology	Lithology affects the permeability of the materials, both surface and sub-surface and thus contributes to the stability.	Geological Map, Indonesian Geological Agency	1:100,000	Vector
Distance to fault	Active faults increase the probability of mass movement in the area around the faults.	Geological Map, Indonesian Geological Agency	1:100,000	Vector
Soil	Different soil characteristics consist of different permeability which influences the slope stability.	Indonesian Center for Agricultural Land Resources Research & Development (ICALRD), 2016	1:50,000	Vector
Environmental				
Annual rainfall	The surface and groundwater dynamics due to rainfall intensity may disturb the slope stability	CHIRPS, 2013 - 2022	0.05°	Raster

Topographical Factors	Description	Source	Resolution/ Scale	Type
NDVI	NDVI corresponds to vegetation coverage. It is associated with soil strength.	Sentinel-2, 2021	10 m (Band 4 and 8)	Raster
Anthropological				
LULC	The pattern of LULC characterises the degree of soil degradation, which further influences the slope stability.	Sentinel-2, 2021	10 m (Band 2, 3, 4, and 8) 20 m (Band 5, 6, 7, 8A, 11, and 12)	Raster
Distance to road	The construction of the road alters the natural landscape and thus impacts on the slope stability.	Indonesian Geospatial information Agency, 2019	1:25,000	Vector

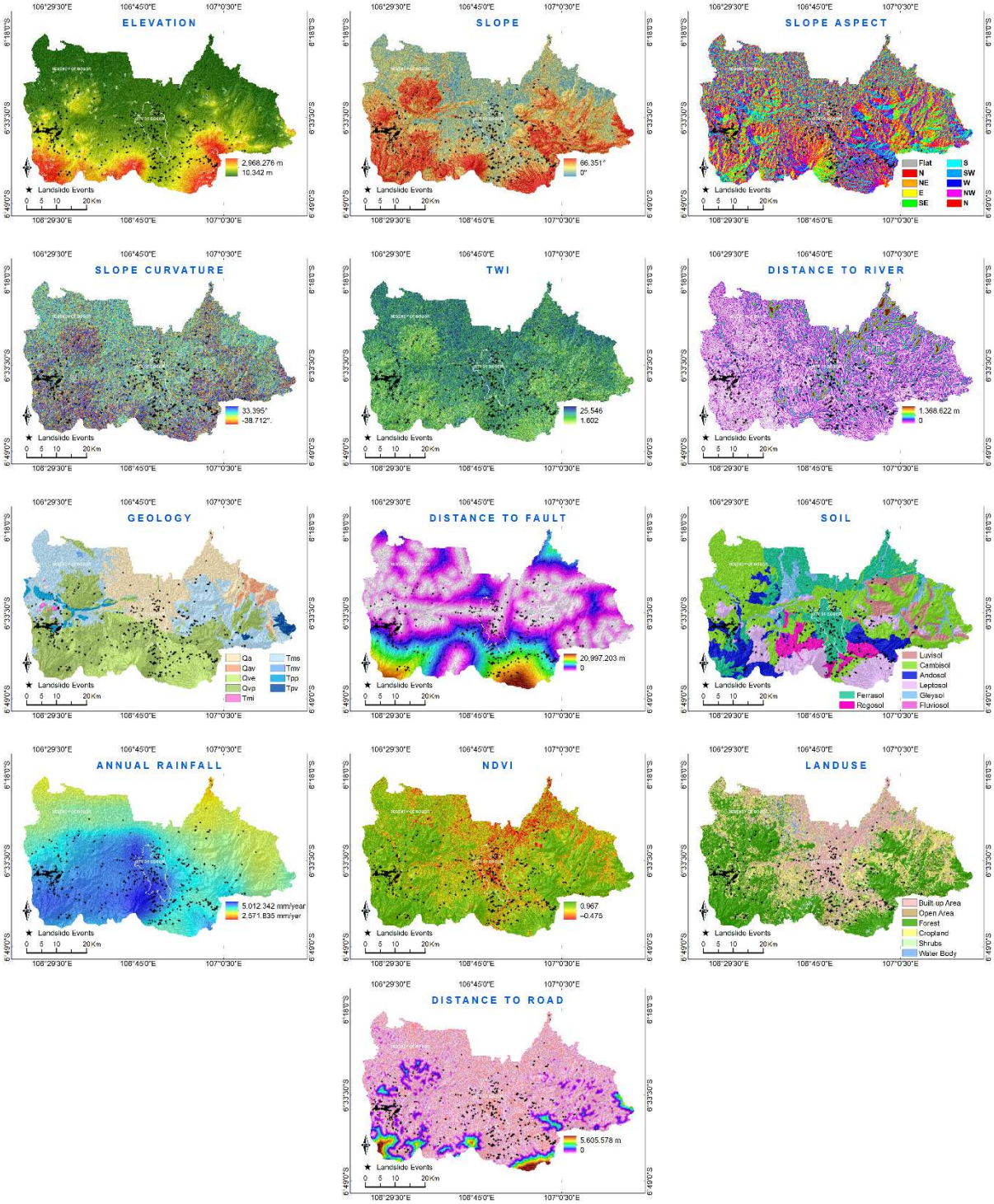


Fig. 2 Landslide Influencing Factor Map

E. Frequency Ratio

Frequency ratio (FR) is the landslide occurrence ratio with the ratio of each class in influencing factors [24]. If the ratio is more significant than 1.0, then the relationship between landslide events and the influencing factors is stronger [27]–[29]. The frequency ratio can be calculated by using equation (1).

$$FR = \frac{\frac{ls(nm)}{\sum ls(m)}}{\frac{grid(nm)}{\sum grid(m)}} \quad (1)$$

where FR is the frequency ratio, $ls(nm)$ is landslide pixel number within the class n of m parameter, $\sum ls(m)$ is the total landslide pixel number of m parameter, $grid(nm)$ is pixel number within class n of m parameter, $\sum grid(m)$ is the total pixel number of m parameter [24], [27], [30], [31]. After calculating the frequency ratio value, the frequency ratio value is normalized into the range 0-1 using min-max normalization [27], [30]. The normalized value is the value of the pixel to be computed.

F. Logistic Regression

One of the classification algorithms used for landslide prediction modeling is logistic regression [22], [32]–[35]. The logistic function for multivariable logistic regression is expressed by equation (2).

$$P(X) = \frac{z}{1+e^z} \quad (2)$$

$P(X)$ is the predicted occurrence of landslides with values in the range $[0, 1]$, and z is a linear fitting function and is expressed in the equation:

$$z = b_0 + b_1X_1 + b_2X_2 + \cdots + b_mX_m \quad (3)$$

where $X = \{X_1, X_2, \ldots, X_m\}$ is predictor (landslide influencing factors) and $\{b_1, b_2, \ldots, b_m\}$ is the coefficient that correlates with predictors.

In modeling landslide prediction using logistic regression, it is necessary to check the multicollinearity of all factors that cause landslides [36]. This stage is intended to see the relationship between two or more factors. If there is a high correlation, the standard error of the coefficients will increase. VIF and TOL are used to assess multicollinearity. The formula for calculating VIF and TOL is by equation (4) [36], [37].

$$VIF_i = \frac{1}{1-R_i^2} = \frac{1}{Tolerance} \quad (4)$$

where R_i is the coefficient of the regression of the independent factor i. The value of $VIF > 10$ indicates the presence of significant multicollinearity and must be eliminated [23], [27], [38], [39].

G. Validation

In this study, the receiver operating characteristics (ROC) curve analysis was performed for model validation [22], [24]. The ROC curve is the curve and the area under the curve (AUC) value. The minimum AUC value is 0.5, and the maximum value of the AUC value is 1, which means the perfect model. The AUC value can be classified into five classes, namely 0.9 – 1 (excellent), 0.8 – 0.9 (very good), 0.7 – 0.8 (good), 0.6 – 0.7 (moderate), and 0.5 – 0.6 (bad) [31]. The formula of AUC of the frequency ratio model is defined by equation (5) [24].

$$AUC = \sum_{i=0}^{n}(x_i - x_{i-1})y_i - \left[\frac{(x_i - x_{i-1})(y_i - y_{i-1})}{2}\right] \quad (5)$$

where x_i is the percentage of the area and y_i is the area of the landslide.

H. Landslide Susceptibility Index (LSI)

The landslide susceptibility index (LSI) of the frequency ratio model is the result of the sum of the frequency ratio values of all parameters on the factors that cause landslides. LSI of the frequency ratio model can be calculated using equation 6 [24], [39].

$$LSI = \sum_{i=1}^{n} FR_i \quad (6)$$

where n is the number of landslide causal factors and FR_i is the frequency ratio of the pixel of landslide causal factor. The LSI of the logistic regression model is generated from equation 3. The higher the LSI value, the higher the susceptibility to landslides, and vice versa [39]. After obtaining the LSI in the study area, LSM is produced by applying classification to the LSI value.

III. Result and Discussions

The landslide point data is divided into two data sets. The first data set is called training data, with a total of 70% or 575 points. The second data set is called validation data, with 30% or 247 points as validation. Training data is used to create a model, and LSI, data validation is used to validate the landslide susceptibility map. The landslide influencing factors have a raster format with the exact resolution and number of pixels. The pixel resolution is 25mx25m, and the number of pixels used for the calculation is 4,934,899 pixels.

A. Frequency Ratio

The frequency ratio (FR) determines the correlation between landslide points and the factors that cause landslides. If a class has an FR value > 1, then the class has a high correlation with landslide events, whereas if a class has an FR value <1, then this class has a low correlation with landslide events [30]. The frequency ratio values for each class of the thirteen factors that cause landslides are shown in Table 2. From the table, the 600-900m class in the elevation parameter has the most incredible FR value of 5.9144 and is hugely different from the other classes. Based on these FR values, class (600-900] m has a high probability of landslide occurrence. The same thing also happens for class (4000.0, 4500.0] in the annual rainfall parameter, which has FR 3.3892 and has a huge difference with the other classes in the annual rainfall parameter. For the slope parameter, classes with intervals (8.0, 17.0] and (17.0, 24.0] have the highest FR values and > 1, namely 2.25439 and 2.1420, respectively.

TABLE II
FREQUENCY RATIO (FR) VALUE FOR EACH LANDSLIDE (LS) INFLUENCING FACTORS

Inventory parameter and interval	LS pixel number	Pixel in domain	%LS (x)	%class (y)	FR (x/y)
Elevation (m)					
<300	145	2,607,114	0.2522	0.5283	0.4773
(300.0, 600.0]	245	1,148,894	0.4261	0.2328	1.8302
(600.0, 900.0]	388	563,030	0.6748	0.1141	5.9144
(900.0, 1200.0]	30	322,891	0.0522	0.0654	0.7974
(1200.0, 1500.0]	14	183,556	0.0243	0.0372	0.6546
>1.500	0	109,414	-	0.0222	-
Slope (°)					
<2	27	494,639	0.0470	0.1002	0.4685
(2.0, 5.0]	81	1,204,885	0.1409	0.2442	0.5770
(5.0, 8.0]	113	894,031	0.1965	0.1812	1.0848
(8.0, 17.0]	412	1,389,978	0.7165	0.2817	2.5439
(17.0, 24.0]	127	508,857	0.2209	0.1031	2.1420
(24.0, 33.0]	52	342,387	0.0904	0.0694	1.3035
>33	10	100,122	0.0174	0.0203	0.8572
Aspect					
North (0°to 22.5°)	69	428,725	0.1200	0.0869	1.3813
Northeast (22.5° to 67.5°)	136	745,289	0.2365	0.1510	1.5661
East (67.5° to 112.5°)	97	597,497	0.1687	0.1211	1.3933
Southeast (112.5° to 157.5°)	70	501,919	0.1217	0.1017	1.1969
South (157.5° to 202.5°)	69	436,802	0.1200	0.0885	1.3557
Southwest (202.5° to 247.5°)	99	498,138	0.1722	0.1009	1.7057
West (247.5° to 292.5°)	95	608,332	0.1652	0.1233	1.3403
Northwest (292.5° to 337.5°)	117	709,101	0.2035	0.1437	1.4161
North (337.5° to 360°)	70	409,096	0.1217	0.0829	1.4685
Curvature					
Concave	364	2,459,198	0.6330	0.4983	1.2703
Flat	0	7,070	-	0.0014	-
Convex	458	2,468,631	0.7965	0.5002	1.5923
TWI					
<4.32	123	534,872	0.2139	0.1084	1.9736
(4.32, 4.88]	149	659,095	0.2591	0.1336	1.9402
(4.88, 5.26]	135	596,991	0.2348	0.1210	1.9408
(5.26, 5.64]	131	628,774	0.2278	0.1274	1.7881
(5.64, 6.77]	189	1,372,202	0.3287	0.2781	1.1821
(6.77, 7.99]	51	597,658	0.0887	0.1211	0.7324
>7.99	44	545,307	0.0765	0.1105	0.6925
Distance to river (m)					
<25	197	1,173,189	0.3426	0.2377	1.4411
(25.0, 50.0]	168	634,621	0.2922	0.1286	2.2720
(50.0, 75.0]	153	651,609	0.2661	0.1320	2.0152
(75.0, 100.0]	81	472,509	0.1409	0.0957	1.4712
(100.0, 125.0]	71	470,968	0.1235	0.0954	1.2938
(125.0, 150.0]	46	285,326	0.0800	0.0578	1.3837
(150.0, 175.0]	33	223,681	0.0574	0.0453	1.2662
(175.0, 200.0]	18	204,867	0.0313	0.0415	0.7541
>200	55	818,129	0.0957	0.1658	0.5770
Lithology					
Tpp - Sediment Pyroclastic (tuffaceous sandstone, conglomerate)	59	97,982	0.1026	0.0199	5.1679
Qve - Quartenary Extrusive Rock (young volcanic deposits: sandy pumiceous tuff, lava, andesitic lahar)	121	733,604	0.2104	0.1487	1.4156
Qvp - Older Volcanic Deposits (breccia, lahar, tuff, tuffaceous sandstone, polymict)	493	1,526,226	0.8574	0.3093	2.7723
Tmi - Intrusive Diorite-Dacite-Andesite	5	16,071	0.0087	0.0033	2.6702

Inventory parameter and interval	LS pixel number	Pixel in domain	%LS (x)	%class (y)	FR (x/y)
Qav - Older Alluvium, polymict, conglomerare-sandstone	0	97,544	-	0.0198	-
Tms - Sedimentary Deposits (sandstone, claystone, narls, limestone)	86	1,360,460	0.1496	0.2757	0.5425
Tpv - Intrusif mangerite-andesite-basalt	0	54,209	-	0.0110	-
Qa - Alluvium Fans, Fluvial/ Flood Plain Deposits	58	1,018,945	0.1009	0.2065	0.4885
Tmv - Neogen Volcanic Deposits (lava, breccia)	0	5,726	-	0.0012	-
Water	0	24,132	-	0.0049	-
Distance to fault (m)					
<1000	98	1,265,888	0.1704	0.2565	0.6644
(1000.0, 2000.0]	120	949,224	0.2087	0.1923	1.0850
(2000.0, 3000.0]	166	646,340	0.2887	0.1310	2.2042
(3000.0, 4000.0]	113	434,289	0.1965	0.0880	2.2331
(4000.0, 5000.0]	125	314,784	0.2174	0.0638	3.4081
(5000.0, 10000.0]	130	836,905	0.2261	0.1696	1.3331
(10000.0, 15000.0]	64	364,295	0.1113	0.0738	1.5078
>15000	6	123,174	0.0104	0.0250	0.4181
Soil					
Regosol	39	261,221	0.0678	0.0529	1.2813
Latosol	182	990,296	0.3165	0.2007	1.5773
Aluvial	1	2,114	0.0017	0.0004	4.0598
Gleisol	20	310,838	0.0348	0.0630	0.5522
Litosol	49	629,374	0.0852	0.1275	0.6682
Andosol	367	515,120	0.6383	0.1044	6.1146
Kambisol	151	1,916,851	0.2626	0.3884	0.6761
Mediteran	13	292,102	0.0226	0.0592	0.3820
Water	0	16,983	-	0.0034	-
Annual rainfall					
<3000	2	59,858	0.0035	0.0121	0.2868
(3000.0, 3500.0]	15	1,157,830	0.0261	0.2346	0.1112
(3500.0, 4000.0]	144	1,581,314	0.2504	0.3204	0.7815
(4000.0, 4500.0]	541	1,369,974	0.9409	0.2776	3.3892
(4500.0, 5000.0]	120	762,498	0.2087	0.1545	1.3507
>5000	0	3,425	-	0.0007	-
NDVI					
<0.1	8	72,617	0.0139	0.0147	0.9455
(0.1, 0.2]	23	168,585	0.0400	0.0342	1.1709
(0.2, 0.3]	31	223,063	0.0539	0.0452	1.1927
(0.3, 0.4]	54	277,186	0.0939	0.0562	1.6720
(0.4, 0.5]	56	327,225	0.0974	0.0663	1.4688
(0.5, 0.6]	83	412,497	0.1443	0.0836	1.7269
(0.6, 0.7]	139	588,860	0.2417	0.1193	2.0259
(0.7, 0.8]	238	1,054,204	0.4139	0.2136	1.9376
>0.8	190	1,810,662	0.3304	0.3669	0.9006
LULC					
Forest	281	2,143,169	0.4887	0.4343	1.1253
Bulit-up area	262	1,397,554	0.4557	0.2832	1.609
Cropland	199	957,734	0.3461	0.1941	1.7833
Shrubs	66	265,643	0.1148	0.0538	2.1323
Open area	14	87,219	0.0243	0.0177	1.3776
Waterbody	0	83,580	-	0.0169	-
Distance to road (m)					
<50	507	1,645,159	0.8817	0.3334	2.6449
(50.0, 100.0]	146	873,267	0.2539	0.1770	1.4349
(100.0, 150.0]	82	619,283	0.1426	0.1255	1.1364
(150.0, 200.0]	35	383,755	0.0609	0.0778	0.7828
(200.0, 250.0]	18	275,720	0.0313	0.0559	0.5603
(250.0, 300.0]	11	162,044	0.0191	0.0328	0.5826
(300.0, 350.0]	5	134,163	0.0087	0.0272	0.3199
(350.0, 400.0]	5	90,742	0.0087	0.0184	0.4729
>400	13	750,766	0.0226	0.1521	0.1486

In aspect factor, all classes have FR>1 values, indicating a high probability of landslides occurring in these classes based on the aspect parameter. For the curvature factor, the occurrence of landslides in the flat class does not exist, and the FR value is 0. This indicates that the probability of landslides occurring in that class is 0. In the TWI factor, the TWI value is inversely proportional to the FR value, this means that the chance of landslides decreases when the TWI

value is high, and vice versa. In the distance to road parameter, the FR value for classes below 200 m has FR > 1, whereas, for road distance> 200 m, the FR value tends to decrease and <1. This indicates that the occurrence of landslides tends to decrease in areas far from the road. The same thing also happened to the distance to river parameter, the distance to river < 175 m class has a value of FR> 1, and the distance to river class > 175 m has a FR value <1, which indicates a high probability of landslides in the area near the river.

Lithology class Tpp - Sediment Pyroclastic (tuffaceous sandstone, conglomerate), Qve - Quartenary Extrusive Rock (young volcanic deposits: sandy pumiceous tuff, lava, andesitic lahar), Qvp - Older Volcanic Deposits (breccia, lava, tuff, tuffaceous), Tmi - Intrusive Diorite-Dacite-Andesite has an FR value >1 and has a huge difference with other classes. Distance to fault in the class between 1000 and 15000 m has a value of FR>1, while for distance to fault <1000 m, the value of FR<1. FR>1 for the soil parameter occurs in the Andosol, alluvial, latosol, and regosol classes, while the Mediterranean, cambisol, litosol, gleisol, and water classes have FR<1. However, the Andosol class has the highest FR values and very huge differences from the other classes. From the calculation of the FR value, it is also possible to obtain a high probability of landslides for the LULC parameter occurring in the class. In the water body for the LULC parameter, the probability of landslides is zero.

B. Logistic Regression

Logistic regression (LR) analysis using scikit sklearn tool [40]. To perform logistic regression analysis, each data consists of independent variables and dependent variables. If the frequency ratio analysis only uses landslide points and outside the landslide point area is considered a non-slide point, then in the logistic regression analysis, the landslide inventory consists of landslide and non-landslide samples with a ratio of 1:1 [28], [41]–[43] and non-landslide points are selected at a minimum distance of 1000 m from the landslide point.

The landslide points used as training data were 575 points and 247 landslide data were used as validation data. For non-landslide data, there are 575 training data and 247 validation data. Thus, the total data is 1,150 training data and 494 validation data. For the values of landslides influencing factor, all data is normalized using min-max normalization, as shown as Fig. 3.

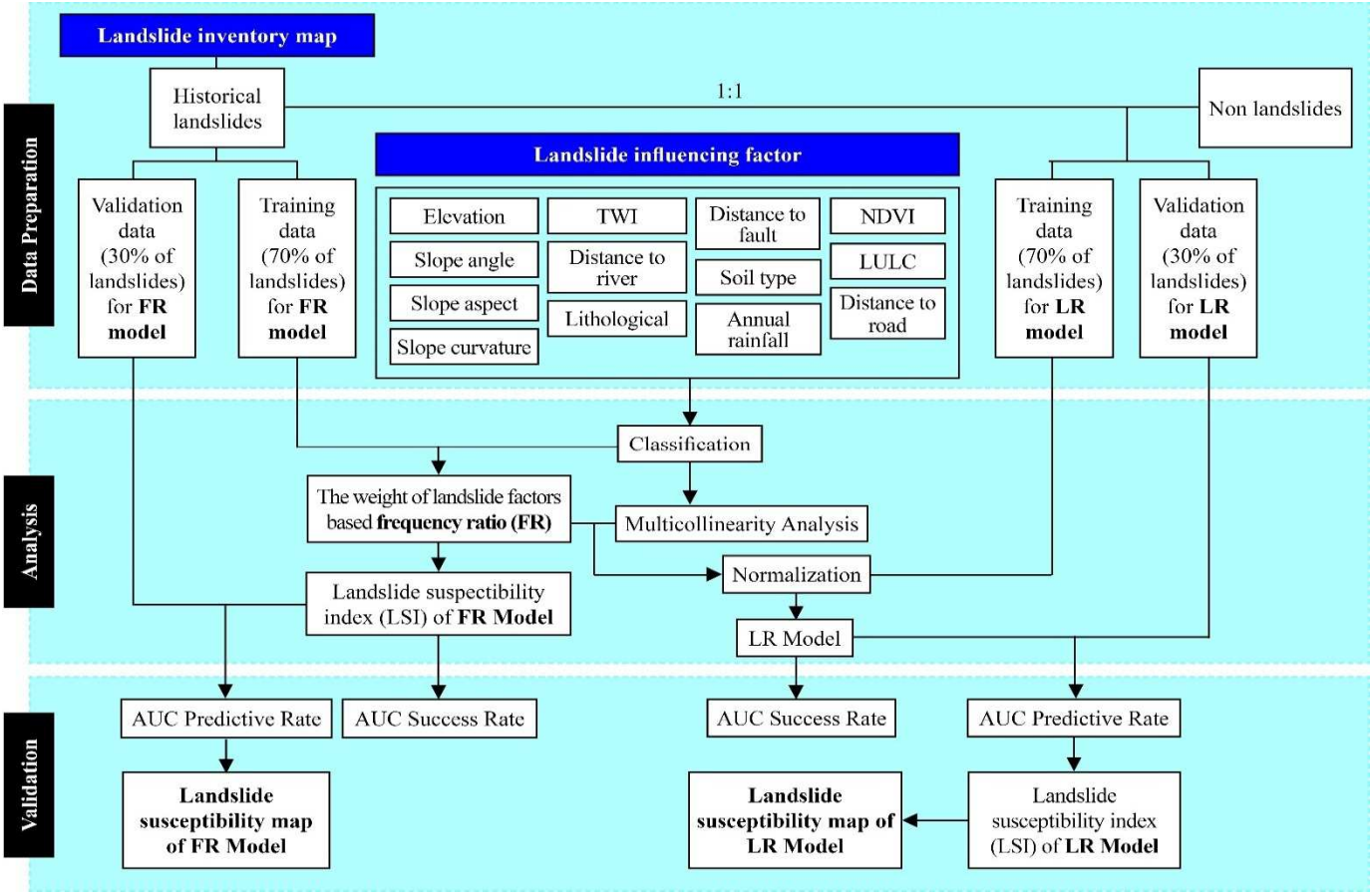


Fig. 3 Flowchart of the research methodology

VIF and TOL are used to assess multicollinearity between predictor variables. The results show that the identified landslide influencing factors are not multicollinear listed in Table 3. The LSI for the LR model is generated by equation (7).

$$LSI = \frac{z}{1+e^{z}} \quad (7)$$

Where:

$$z = -5.9646 + 1.5261X_1 + 1.0631X_2 + 0.5877X_3 + -0.4679X_4 + 0.5219X_5 + 1.3795X_6 + 0.7046X_7 + -0.1146X_8 + 0.5887X_9 + 1.9238X_{10} + 0.7553X_{11} + 1.7053X_{12} + 1.4325X_{13} \quad (8)$$

The value $X = \{X_1, X_2, \ldots, X_{13}\}$ is the landslide influencing factors. LSI values for LR models are in the range 0.0027-0.9963. For LSI LR values that are close to 1, the chance of landslides is higher, for LSI LR values close to 0, the chances of landslides are lower.

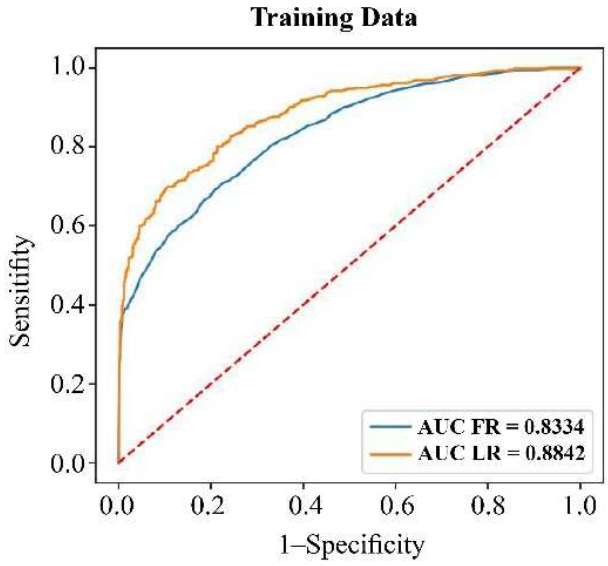
TABLE III
MULTICOLLINEARITY ANALYSIS USING TOLERANCE AND VIF FOR LANDSLIDE INFLUENCING FACTORS

Independent Variables	VIF	Tolerance
Elevation	1.7668	0.566
Lithology	1.7084	0.5853
TWI	1.4794	0.676
Slope	1.4534	0.6881
Distance to Fault	1.3436	0.7443
NDVI	1.3166	0.7595
Annual rainfall	1.2955	0.7719
Slope Curvature	1.2952	0.772
Soil	1.2561	0.7961
LULC	1.224	0.817
Distance to River	1.0757	0.9296
Distance to Road	1.0712	0.9335
Slope Aspect	1.0415	0.9602

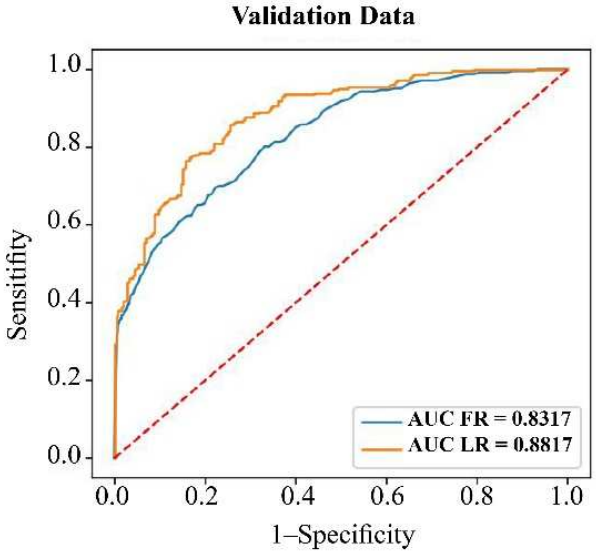
The Landslide Susceptible Map (LSM) for the FR and LR models is built by dividing the LSI into 5 classes using the natural break number [22], [39], [42], [44], [45]. The five interval classes are very low, low, moderate, high, and very high. From Table 4, the LSM from modeling using FR consists of very low (7.8133-15.2421), low (15.2421-18.4773), moderate (18.4773-21.9520), high (21.9520-26.5052), and very high (26.5052-38.3674). The LSM from modeling using LR consist of very low (0.0027-0.1542), low (0.1542-0.3136), moderate (0.3136-0.5001), high (0.5001-0.7139), and very high (0.7139-0.9937).

The LSM of the FR model, there are 24.64% of the total area in the very low class, 31.09% is in the low class, 22.84% is in the moderate class, 16.18% is in the high class, 5.26% is in the very high class. Whereas for the LR model, 34.69% of the total area is in the very low class, 26.13% is in the low class, 17.36% is in the moderate class, 12.68% is in the high class, 9.14% is in the very high class. For the FR model, the most area is low, while for the LR model, the most area is very low. However, both models show that the majority of the area is covered by low and very low LSM classes, while the high and very high classes cover a smaller number of areas.

Validation of the accuracy of this research model using ROC analysis. ROC is one of the methods used to determine the accuracy of the prediction model used with the frequency ratio and logistic regression. The ROC AUC curve generated with the frequency ratio and logistic regression model is shown in Fig. 4. The AUC value of frequency ratio and logistic regression model for data validation are 0.8317 and 0.8817, respectively. It indicates that the LSI calculation using both models produces a good model.



(a)



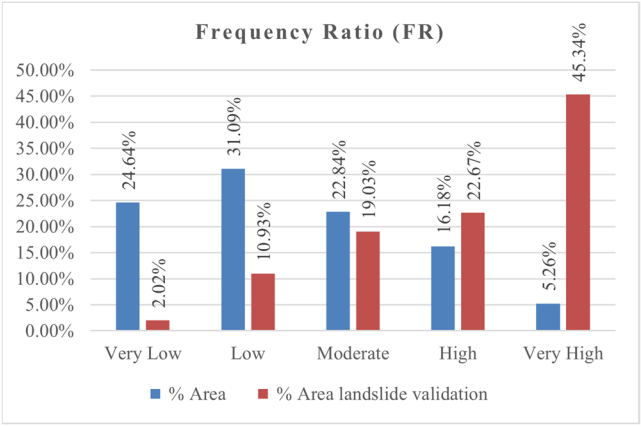
(b)

Fig. 4 Area under (AUC) of the ROC curve (a) Success rate of training data and (b) prediction rate of validation data diagram showing percentage of study area classified as susceptible (x-axis) in cumulative percent of landslide events (y-axis).

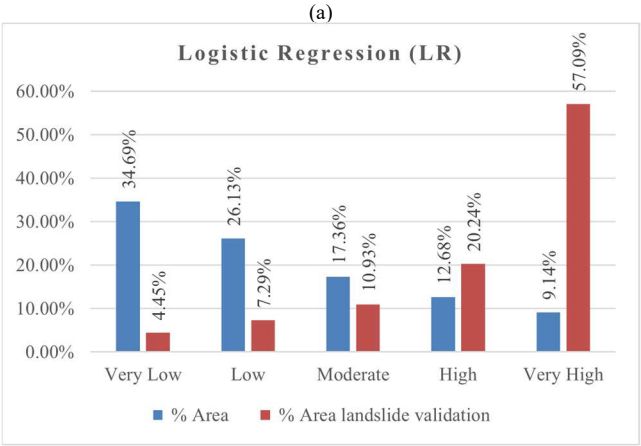
Another way to validate the LR and FR models is also done by validating validation points with LSM, as shown in Table 4 and Fig. 5. For the FR model, 68.01% or 168 landslide points are in the high and very high classes, and only 12.95% are in the low and very low classes. Meanwhile, for the LR model, 77.33% or 191 landslide points were in the high and very high classes, and only 11.74% were in the low and very low classes.

TABLE IV
THE CHARACTERISTICS OF SUSCEPTIBILITY CLASSES ON LSM USING FREQUENCY RATIO AND LOGISTIC REGRESSION MODEL

Class	LSI	Area		Landslides validation	
		Pixels	%	Pixels	%
Frequency Ratio					
Very low	7.8133-15.2421	1,215,861	24.64%	5	2.02%
Low	15.2421-18.4773	1,534,092	31.09%	27	10.93%
Moderate	18.4773-21.9520	1,127,202	22.84%	47	19.03%
High	21.9520-26.5052	798,380	16.18%	56	22.67%
Very high	26.5052-38.3674	259,364	5.26%	112	45.34%
		4,934,899		**247**	
Logistic Regression					
Very low	0.0027-0.1542	1,711,808	34.69%	11	4.45%
Low	0.1542-0.3136	1,289,560	26.13%	18	7.29%
Moderate	0.3136-0.5001	856,718	17.36%	27	10.93%
High	0.5001-0.7139	625,533	12.68%	50	20.24%
Very high	0.7139-0.9937	451,280	9.14%	141	57.09%
		4,934,899		**247**	



(a)



(b)

Fig. 5 Percentage of landslide susceptibility classes and percentage of landslides validation on landslide susceptibility map using (a) frequency ratio and (b) logistic regression.

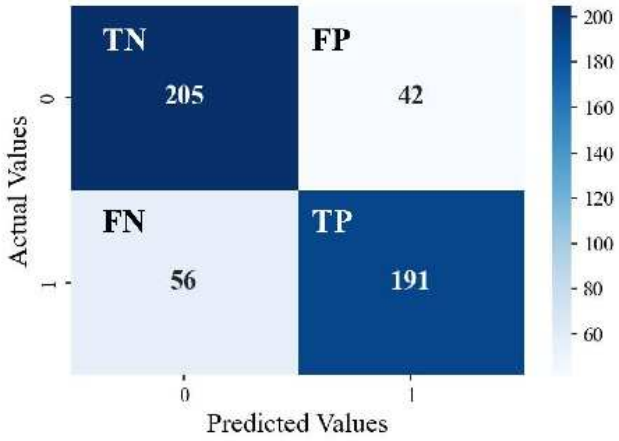


Fig. 6 Confusion matrix of LR model

TABLE V
METRICS OF LOGISTIC REGRESSION MODEL

Metrics	Formula	Score
Sensitivity	$\frac{TP}{TP + FN}$	0.7732
Specificity	$\frac{TN}{TN + FP}$	0.8299
Precision	$\frac{TP}{TP + FP}$	0.8197
F1-score	$\frac{2 * precision * sensitivity}{precision + sensitivity}$	0.7958
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	0.8016

For the LR model, model performance can also be seen from the performance metrics derived from the confusion matrix [22], [33], [39], [42], as shown in Fig. 6. From the confusion matrix, it was found that 247 landslide data were predicted by the LR model to 56 non-landslides (False Negative – FN) and 191 landslides (True Positive – TP). Meanwhile, 247 non-landslide data are predicted by the LR model to 205 non-landslides (True Negative – TN) and 42 landslides (False Positive – FP).

The performance metrics are obtained from the confusion matrix (Table 5), and the values show good results, indicating that the LR model is good for landslide prediction. Accuracy describes how accurately the model is able to classify correctly, formulated by the ratio of correct predictions to all data. The accuracy of the LR model is 0.8016. However, in this LR model, the sensitivity value tends to be lower than the specificity, which indicates that the model tends to be more accurate in predicting non-landslide data.

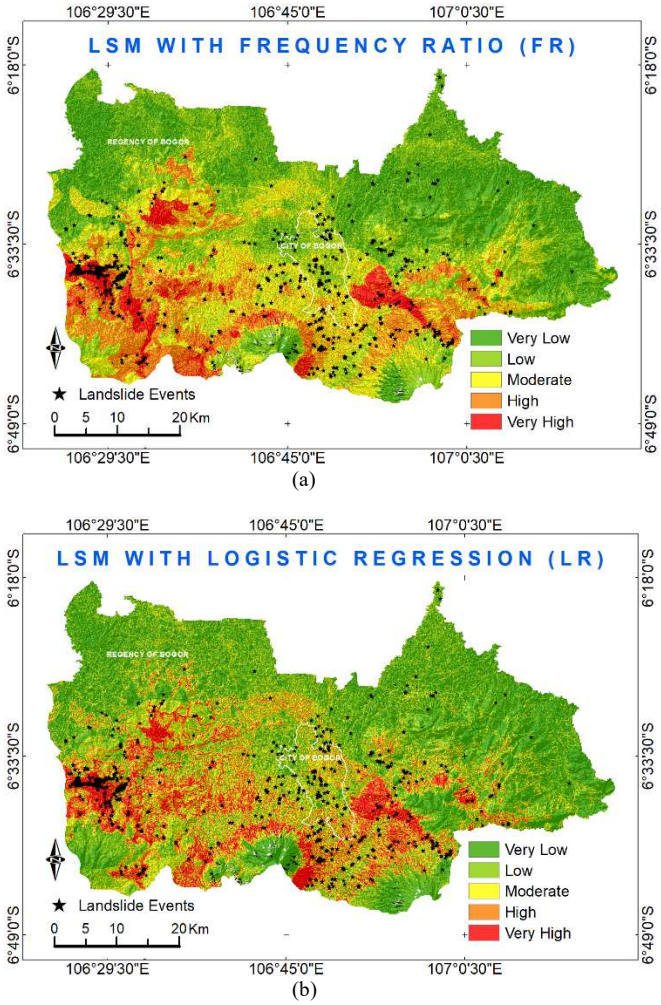


Fig. 7 Landslide susceptibility map using (a) frequency ratio and (b) logistic regression model.

IV. Conclusion

The FR and LR model as a statistical technique was tested in the study area to evaluate the relationship between 13 influencing factors and landslide occurrences. Based on the assessed result, the highest frequency ratios were elevation, lithology, and soil. Those are the basic properties of slope as predisposing factors that control the slope movement. This result led to a proper structural mitigation in the study area regarding rock and soil protection for slope stability, particularly in moderate – to very high susceptibility areas.

The AUC validation processes indicate the model has a high accuracy of the landslides spatial evidence with a score of 0.8334 for the FR model and 0.8842 for the LR model. The model also produces a high prediction of future landslide occurrences, with a score of 0.8317 for the FR model and 0.8817 for the LR model. Both the models produce satisfactory results, although the LR model shows better accuracy overall. However, calculations using the FR model are more straightforward than the LR model. The result determined that landslide inventory greatly affected the accuracy levels. For a comparable outcome, we suggested complete documentation of the historical database, e.g., the relation of different landslide types to each influencing factor.

The presented LSM of Bogor is a medium-scale map for regional-based planning. More advanced and complex models can be applied in the LSM study, but there is yet to be a certain consensus regarding which methods are more superior. Hence, we recommend the FR and LR model as a simple method that can be easily implemented by local government (regency, district, city level).

Acknowledgment

This research was fully supported by BRIN, No: 1/III.6/HK/2023, and under collaboration between The National Research and Innovation Agency (BRIN) and the University of Bina Nusantara.

References

[1] The National Disaster Management Agency, "Statistics of landslide event in Bogor area 2005-2020," *BNPB*, Mar. 2022. https://dibi.bnpb.go.id/ (accessed Mar. 11, 2023).

[2] D. Ozturk and N. Uzel-Gunini, "Investigation of the effects of hybrid modeling approaches, factor standardization, and categorical mapping on the performance of landslide susceptibility mapping in Van, Turkey," *Nat. Hazards*, vol. 114, no. 3, pp. 2571–2604, Dec. 2022, doi:10.1007/s11069-022-05480-y.

[3] Z. Qin, X. Zhou, M. Li, Y. Tong, and H. Luo, "Landslide Susceptibility Mapping Based on Resampling Method and FR-CNN: A Case Study of Changdu," *Land*, vol. 12, no. 6, p. 1213, Jun. 2023, doi:10.3390/land12061213.

[4] W. Wu, S. Guo, and Z. Shao, "Landslide risk evaluation and its causative factors in typical mountain environment of China: a case study of Yunfu City," *Ecol. Indic.*, vol. 154, p. 110821, Oct. 2023, doi:10.1016/j.ecolind.2023.110821.

[5] L. Zhu *et al.*, "Landslide Susceptibility Prediction Modeling Based on Remote Sensing and a Novel Deep Learning Algorithm of a Cascade-Parallel Recurrent Neural Network," *Sensors*, vol. 20, no. 6, p. 1576, Mar. 2020, doi: 10.3390/s20061576.

[6] Center for Volcanology and Geological Hazard Mitigation Geological Agency, "PVMBG Ungkap Pulau Jawa Rawan Longsor," *https://www.krjogja.com/peristiwa/read/492328/pvmbg-ungkap-pulau-jawa-rawan-longsor*, 2022. [Online]. Available: https://www.krjogja.com/nasional/1242457473/-pvmbg-ungkap-pulau-jawa-rawan-longsor

[7] Statistics Indonesia, "Statistical Yearbook of Indonesia 2023," BPS-Statistics Indonesia, 2023. [Online]. Available: https://www.bps.go.id/publication/2023/02/28/18018f9896f09f03580a614b/statistik-indonesia-2023.html

[8] Statistics of Jawa Barat Province, *Jawa Barat Province in Figures 2023*. 2023. [Online]. Available: https://jabar.bps.go.id/publication/2023/02/28/57231a828abbfdd50a21fe31/provinsi-jawa-barat-dalam-angka-2023.html

[9] A. Addis, "GIS-Based Landslide Susceptibility Mapping Using Frequency Ratio and Shannon Entropy Models in Dejen District, Northwestern Ethiopia," *J. Eng.*, vol. 2023, pp. 1–14, Feb. 2023, doi:10.1155/2023/1062388.

[10] G. Berhane *et al.*, "Landslide susceptibility zonation mapping using GIS-based frequency ratio model with multi-class spatial data-sets in the Adwa-Adigrat mountain chains, northern Ethiopia," *J. African Earth Sci.*, vol. 164, p. 103795, Apr. 2020, doi:10.1016/j.jafrearsci.2020.103795.

[11] I. Cantarino, M. A. Carrion, V. Martínez-Ibáñez, and E. Gielen, "Improving Landslide Susceptibility Assessment through Frequency Ratio and Classification Methods—Case Study of Valencia Region (Spain)," *Appl. Sci.*, vol. 13, no. 8, p. 5146, Apr. 2023, doi:10.3390/app13085146.

[12] W. He *et al.*, "Landslide Susceptibility Evaluation of Machine Learning Based on Information Volume and Frequency Ratio: A Case Study of Weixin County, China," *Sensors*, vol. 23, no. 5, p. 2549, Feb. 2023, doi: 10.3390/s23052549.

[13] B. Li, N. Wang, and J. Chen, "GIS-Based Landslide Susceptibility Mapping Using Information, Frequency Ratio, and Artificial Neural Network Methods in Qinghai Province, Northwestern China," *Adv. Civ. Eng.*, vol. 2021, pp. 1–14, Jun. 2021, doi: 10.1155/2021/4758062.

[14] Y. Liu, A. Yuan, Z. Bai, and J. Zhu, "GIS-based landslide susceptibility mapping using frequency ratio and index of entropy models for She County of Anhui Province, China," *Appl. Rheol.*, vol.

32, no. 1, pp. 22–33, Jun. 2022, doi: 10.1515/arh-2022-0122.
[15] I. Sonker, J. N. Tripathi, and Swarnim, "Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio method in Sikkim Himalaya," *Quat. Sci. Adv.*, vol. 8, p. 100067, Oct. 2022, doi: 10.1016/j.qsa.2022.100067.
[16] S. Panchal and A. K. Shrivastava, "Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India," *Ain Shams Eng. J.*, vol. 13, no. 3, p. 101626, May 2022, doi: 10.1016/J.ASEJ.2021.10.021.
[17] A. H. Alsabhan *et al.*, "Landslide susceptibility assessment in the Himalayan range based along Kasauli – Parwanoo road corridor using weight of evidence, information value, and frequency ratio," *J. King Saud Univ. - Sci.*, vol. 34, no. 2, p. 101759, Feb. 2022, doi:10.1016/j.jksus.2021.101759.
[18] A. Es-smairi, B. Elmoutchou, R. A. Mir, A. El Ouazani Touhami, and M. Namous, "Delineation of landslide susceptible zones using Frequency Ratio (FR) and Shannon Entropy (SE) models in northern Rif, Morocco," *Geosystems and Geoenvironment*, vol. 2, no. 4, p. 100195, Nov. 2023, doi: 10.1016/j.geogeo.2023.100195.
[19] A. Ozdemir, "A Comparative Study of the Frequency Ratio, Analytical Hierarchy Process, Artificial Neural Networks and Fuzzy Logic Methods for Landslide Susceptibility Mapping: Taşkent (Konya), Turkey," *Geotech. Geol. Eng.*, vol. 38, no. 4, pp. 4129–4157, Aug. 2020, doi: 10.1007/s10706-020-01284-8.
[20] B. Hafsa, M. S. Chowdhury, and M. N. Rahman, "Landslide susceptibility mapping of Rangamati District of Bangladesh using statistical and machine intelligence model," *Arab. J. Geosci.*, vol. 15, no. 15, p. 1367, Aug. 2022, doi: 10.1007/s12517-022-10607-3.
[21] A. R. Rasyid, N. P. Bhandary, and R. Yatabe, "Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia," *Geoenvironmental Disasters*, vol. 3, no. 1, p. 19, Dec. 2016, doi:10.1186/s40677-016-0053-x.
[22] R.-X. Tang, E.-C. Yan, T. Wen, X.-M. Yin, and W. Tang, "Comparison of Logistic Regression, Information Value, and Comprehensive Evaluating Model for Landslide Susceptibility Mapping," *Sustainability*, vol. 13, no. 7, p. 3803, Mar. 2021, doi:10.3390/su13073803.
[23] P. Kainthura and N. Sharma, "Hybrid machine learning approach for landslide prediction, Uttarakhand, India," *Sci. Rep.*, vol. 12, no. 1, p. 20101, Nov. 2022, doi: 10.1038/s41598-022-22814-9.
[24] F. E. S. Silalahi, Pamela, Y. Arifianti, and F. Hidayat, "Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia," *Geosci. Lett.*, vol. 6, no. 1, p. 10, Dec. 2019, doi:10.1186/s40562-019-0140-4.
[25] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.
[26] D. N. Melati, Astisiasari, and Trinugroho, "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, Dec. 2022, pp. 73–78. doi:10.1109/AGERS56232.2022.10093267.
[27] Y. Xing, J. Yue, Z. Guo, Y. Chen, J. Hu, and A. Travé, "Large-Scale Landslide Susceptibility Mapping Using an Integrated Machine Learning Model: A Case Study in the Lvliang Mountains of China," *Front. Earth Sci.*, vol. 9, Aug. 2021, doi: 10.3389/feart.2021.722491.
[28] M. Sheng, J. Zhou, X. Chen, Y. Teng, A. Hong, and G. Liu, "Landslide Susceptibility Prediction Based on Frequency Ratio Method and C5.0 Decision Tree Model," *Front. Earth Sci.*, vol. 10, May 2022, doi:10.3389/feart.2022.918386.
[29] H. Shu, Z. Guo, S. Qi, D. Song, H. Pourghasemi, and J. Ma, "Integrating Landslide Typology with Weighted Frequency Ratio Model for Landslide Susceptibility Mapping: A Case Study from Lanzhou City of Northwestern China," *Remote Sens.*, vol. 13, no. 18, p. 3623, Sep. 2021, doi: 10.3390/rs13183623.
[30] Y. Wang, D. Sun, H. Wen, H. Zhang, and F. Zhang, "Comparison of Random Forest Model and Frequency Ratio Model for Landslide Susceptibility Mapping (LSM) in Yunyang County (Chongqing, China)," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, p. 4206, Jun. 2020, doi: 10.3390/ijerph17124206.
[31] S. Sukristiyanti, K. Wikantika, I. A. Sadisun, L. F. Yayusman, A. Tohari, and M. H. Zaenal Putra, "Evaluation of Parameter Selection in the Bivariate Statistical-based Landslide Susceptibility Modeling (Case Study: the Citarik Sub-watershed, Indonesia)," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 1, p. 244, Jan. 2022, doi:10.18517/ijaseit.12.1.14737.
[32] A. Wubalem, "Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia," *Geoenvironmental Disasters*, vol. 8, no. 1, p. 1, Dec. 2021, doi:10.1186/s40677-020-00170-y.
[33] H. Wang, L. Zhang, K. Yin, H. Luo, and J. Li, "Landslide identification using machine learning," *Geosci. Front.*, vol. 12, no. 1, pp. 351–364, Jan. 2021, doi: 10.1016/j.gsf.2020.02.012.
[34] T. Zhang *et al.*, "Evaluation of different machine learning models and novel deep learning-based algorithm for landslide susceptibility mapping," *Geosci. Lett.*, vol. 9, no. 1, p. 26, Dec. 2022, doi:10.1186/s40562-022-00236-9.
[35] X. Zhou, W. Wu, Y. Qin, and X. Fu, "Geoinformation-based landslide susceptibility mapping in subtropical area," *Sci. Rep.*, vol. 11, no. 1, p. 24325, Dec. 2021, doi: 10.1038/s41598-021-03743-5.
[36] X. Chen and W. Chen, "GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods," *CATENA*, vol. 196, p. 104833, Jan. 2021, doi: 10.1016/j.catena.2020.104833.
[37] L.-L. Liu, C. Yang, F.-M. Huang, and X.-M. Wang, "Landslide susceptibility mapping by attentional factorization machines considering feature interactions," *Geomatics, Nat. Hazards Risk*, vol. 12, no. 1, pp. 1837–1861, Jan. 2021, doi:10.1080/19475705.2021.1950217.
[38] 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, p. 1451, Jan. 2022, doi: 10.1038/s41598-022-05364-y.
[39] Q. B. Pham *et al.*, "A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping," *Geomatics, Nat. Hazards Risk*, vol. 12, no. 1, pp. 1741–1777, Jan. 2021, doi:10.1080/19475705.2021.1944330.
[40] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011, [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html
[41] K. Gaidzik and M. T. Ramírez-Herrera, "The importance of input data on landslide susceptibility mapping," *Sci. Rep.*, vol. 11, no. 1, p. 19334, Sep. 2021, doi: 10.1038/s41598-021-98830-y.
[42] A. Merghadi *et al.*, "Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance," *Earth-Science Rev.*, vol. 207, p. 103225, Aug. 2020, doi: 10.1016/j.earscirev.2020.103225.
[43] W. Wu *et al.*, "A Data-Driven Model on Google Earth Engine for Landslide Susceptibility Assessment in the Hengduan Mountains, the Qinghai–Tibetan Plateau," *Remote Sens.*, vol. 14, no. 18, p. 4662, Sep. 2022, doi: 10.3390/rs14184662.
[44] J. Dou *et al.*, "Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan," *Landslides*, vol. 17, no. 3, pp. 641–658, Mar. 2020, doi: 10.1007/s10346-019-01286-5.
[45] A. Wubalem and M. Meten, "Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia," *SN Appl. Sci.*, vol. 2, no. 5, p. 807, May 2020, doi: 10.1007/s42452-020-2563-0.