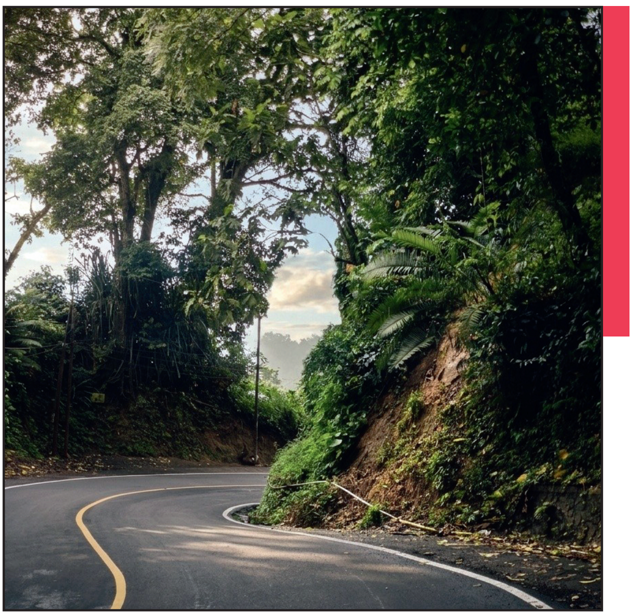
Acta geographica Slovenica, 66-1, 2026, 179–196

MACHINE LEARNING-BASED LANDSLIDE SUSCEPTIBILITY ASSESSMENT FOR SHALLOW LANDSLIDE MITIGATION IN THE GUMITIR MOUNTAIN AREA, INDONESIA

Entin Hidayah, Mokhammad Farid Maruf, Siti Norafida Jusoh, Joko Widodo, Afif Amiluddin


JAGAT ADI SAMUDRA

Potential landslides on the prone path of Mount Gumitir, Jember.

DOI: https://doi.org/10.3986/AGS.14449
UDC: 551.435.62(594.59)

Entin Hidayah[1], Mokhammad Farid Maruf[1], Siti Norafida Jusoh[2], Joko Widodo[3], Afif Amiluddin[4]

Machine learning-based landslide susceptibility assessment for shallow landslide mitigation in the Gumitir Mountain Area, Indonesia

ABSTRACT: Recurrent rainfall-induced landslides in Gumitir, East Java, threaten the Jember–Banyuwangi route. This study aims to develop mitigation strategies by analyzing influencing factors and predicting landslide-prone areas using logistic regression and random forest models. A dataset consisting of 96 landslide and 180 non-landslide points was randomly divided into a set for model training (70%) and testing (30%). The random forest model outperformed logistic regression with slope, elevation, flow accumulation, topographic wetness index, normalized difference vegetation index, and aspect identified as the most influential factors. These results support targeted and effective landslide risk mitigation.

KEYWORDS: landslide susceptibility assessment, logistic regression, random forest modeling, geospatial analysis, slope stability, hydrological factors, mountainous area

Ocena plazovitosti na podlagi strojnega učenja za zmanjšanje tveganja nastanka plitvih plazov na območju gore Gumitir v Indoneziji

POVZETEK: Ponavljajoči se plazovi, ki se na območju gore Gumitir na Vzhodni Javi sprožajo zaradi dežja, ogrožajo prometno povezavo med mestoma Jember in Banyuwangi. Namen članka je razviti strategije za zmanjšanje tveganja nastanka plazov na podlagi analize vplivnih dejavnikov in napovedi plazovitih območij z modeli logistične regresije in naključnih gozdov. Podatkovni niz, sestavljen iz 96 lokacij plazov in 180 lokacij brez plazov, je bil naključno razdeljen v množico za učenje (70%) in testiranje (30%) modelov. Z modelom naključnih gozdov so bili doseženi boljši rezultati kot z modelom logistične regresije, pri čemer so bili kot najvplivnejši dejavniki opredeljeni naklon, nadmorska višina, akumulacija toka, indeks topografske vlažnosti, normaliziran diferencialni vegetacijski indeks in ekspozicija. Rezultati podpirajo ciljno usmerjeno in učinkovito zmanjševanje tveganja nastanka plazov.

KLJUČNE BESEDE: ocena plazovitosti, logistična regresija, modeliranje z metodo naključnega gozda, geoprostorska analiza, stabilnost pobočij, hidrološki dejavniki, gorsko območje

The article was submitted for publication on May 14th 2025.
Uredništvo je prejelo prispevek 14. maja 2025.

[1] University of Jember, Department of Civil Engineering, Jember, Indonesia
[2] Universiti Teknologi Malaysia, Faculty of Civil Engineering, Johor Bahru, Johor, Malaysia
[3] Research Center for Geoinformatics, National Research and Innovation Agency (BRIN), Bogor, Indonesia
[4] Universitas Mochammad Sroedji, Civil Engineering Department, Engineering Faculty, Jember, Indonesia

entin.teknik@unej.ac.id (https://orcid.org/0000-0002-1233-6850), farid.teknik@unej.ac.id (https://orcid.org/0000-0002-5715-935X), snorafida@utm.my (https://orcid.org/0000-0002-7132-0883), jecko.emji2@gmail.com (https://orcid.org/0000-0002-0552-5100), afif@umsj.ac.id (https://orcid.org/0009-0000-8600-0922)

1 Introduction

Landslide hazards in tropical mountainous regions have intensified due to extreme rainfall patterns linked to climate change, threatening critical infrastructure such as road networks (Alcántara-Ayala 2025). It is the most significant geological hazard that poses a major threat to life, infrastructure, and environmental stability. A range of factors, such as intense rainfall, slope geometry, geological conditions, weak or collapsible soil materials, groundwater table position, deforestation, and many more, can influence the landslide occurrence. In tropical regions, this is further aggravated by the high-intensity rainfall events that elevate pore water pressure, reduce the shear strength in soil, and hence increase the slope failure probability (Rosly et al. 2023; Sarah et al. 2024). In Gumitir, East Java, Indonesia, many slopes have been explored for farming purposes and modified without comprehensive geotechnical investigation or proper design, hence inducing recurrent landslides along the Jember-Banyuwangi highway that disrupt regional connectivity, endanger lives, and incur high economic costs. The current approach to landslide mitigation is still reactive and not preventive. It focuses on landslide-prone spots only after a landslide has occurred. Addressing landslides at these specific spots provides only temporary solutions without addressing the root causes, leaving problems for the future, as is the case today. Referring to previous mitigation efforts, there is an urgent need to handle landslide occurrences. Hence, it is important to develop future landslide mitigation strategies to reduce adverse impacts and address landslide occurrences based on scientific approaches for landslide susceptibility mapping (Getachew and Meten 2021). To develop a more comprehensive landslide mitigation strategy, a vulnerability assessment is necessary (Abdulwahid and Pradhan 2017). In landslide vulnerability assessments, understanding the spatial causes of landslides and predicting potential landslide locations is crucial. Both geomorphological and environmental factors contribute significantly to landslides, making their investigation essential (Kohno and Higuchi 2023). These factors will naturally have a strong impact on the landslide distribution.

Recent disadvantages and advances in machine learning (ML) and geospatial analysis have enabled data-driven landslide susceptibility assessments in identifying risk zones. While advanced techniques like support vector machines (SVM), k-means, hierarchical clustering, self-organizing maps (SOMs), artificial neural networks (ANNs), and Bayesian networks have been applied in landslide susceptibility studies, their high computational complexity, extensive data requirements, and limited interpretability hinder practical adoption (Le et al. 2024; Teng et al. 2024). Decision trees in landslide prediction struggle with overfitting and parameter sensitivity, leading to unreliable predictions and reduced interpretability as deep trees obscure dominant factors, limiting their practical utility for policymakers (Park et al. 2019; Teng et al. 2024). The multi-layer perceptron (MLP) model requires high computational resources and a large training dataset, and although it offers high accuracy, it is difficult to interpret due to its nature (Teng et al. 2024). Random forest (RF) and logistic regression (LR) methods minimize error and provide reasonable spatial resolutions (Huang, F. et al. 2023). The application of LR is highly beneficial for predicting landslide susceptibility, representing the outcomes based on the influential factors triggering landslides (Huang, F. et al. 2023). Meanwhile, RF has several advantages, including reducing the risk of overfitting (Munawar et al. 2022; Rong et al. 2023). This algorithm can be used for the problems of both regression and classification (Saha et al. 2022; Le et al. 2024) and is effective for determining variable importance by measuring the decrease in model accuracy (Achour and Pourghasemi 2020; Huang, F. et al. 2023; Rong et al. 2023). However, most studies focus on the accuracy of susceptibility mapping without translating results into actionable mitigation frameworks, especially in tropical road networks where slope stability, drainage, and vegetation dynamics interact uniquely. Furthermore, the role of aspect (slope orientation) in landslide initiation remains understudied, as its influence is often conflated with elevation and rainfall patterns. In shallow landslides, south-facing slopes exhibit distinct soil properties (higher silt, slower pore-pressure dissipation), necessitating aspect-specific models for accurate predictions (Guo and Ma 2023).

This study proposes an integrated ML-geospatial framework using the example of landslide risks in Gumitir, East Java, through three primary objectives: (1) identifying dominant geospatial triggers of landslides, including slope, elevation, aspect, and hydrological factors, (2) evaluating the comparative performance of RF and LR models in predicting landslide-prone zones under tropical mountainous conditions, and (3) translating susceptibility mapping results into spatially targeted mitigation strategies to enhance road resilience. The work seeks to determine the relative influence of geospatial variables on landslide susceptibility, assess the robustness of RF and LR in high-rainfall environments, and demonstrate how predictive models can directly inform engineering interventions, such as reinforced drainage systems in high-flow accumulation areas or slope stabilization measures in steep terrain, to reduce risks proactively.

2 Study area

This study focuses on the Gumitir Mountain area, located along the road connecting Jember and Banyuwangi districts (Figure 1). The region lies on the border of both districts and is hereafter referred to as the »Gumitir Mountain Area.« It covers a micro-watershed of approximately 1,814.07 ha, with elevations ranging from 420 to 967 m above sea level.

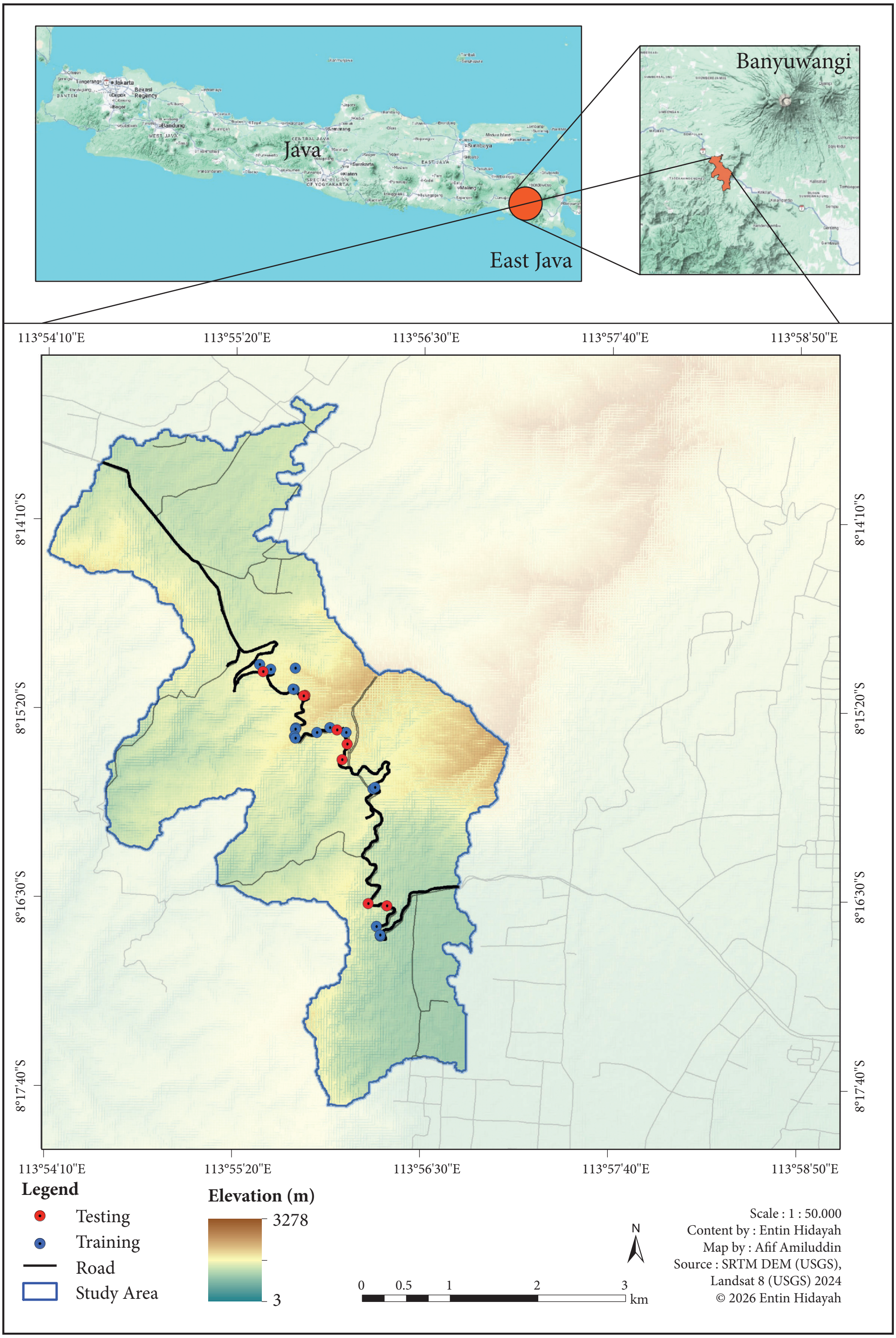
The Gumitir Mountain Area is predominantly mountainous, with land cover consisting mainly of forest and plantations. This region has experienced multiple landslides in the past, particularly along the Jember–Banyuwangi road corridor, where slope failures have disrupted traffic and damaged infrastructure (Figure 2). As illustrated in Figure 2a), the observed failures are predominantly shallow translational soil slides developing within the weathered, clay-rich residual soil layer and occurring in proximity to the road embankment. Figure 2b) further shows damage to road infrastructure, including deformed guardrails and exposed utility cables, indicating the combined effects of slope instability, surface erosion, and localized drainage disturbance along the road corridor.

The concentration of mapped landslides along the road corridor may partly reflect an observational bias, as slope failures in remote and inaccessible valleys are more difficult to detect in the absence of road access. However, the morphological evidence observed adjacent to the road also indicates that road cuttings and localized drainage disturbances have directly contributed to slope instability in the Gumitir Mountain Area. Increasing population pressure has intensified land use in hilly terrain, making slope utilization and infrastructure development unavoidable. Consequently, mapping landslide-prone areas is essential for identifying potential hazards and supporting risk-informed and sustainable development planning.

Geologically, the Gumitir Mountain Area range is entirely underlain by the Batuampar Formation, consisting of medium-grained clastic sandstone deposited in a littoral environment. Prolonged tropical weathering has transformed this formation into thick, clay-rich residual soils with low permeability, resulting in a geologically fragile setting. Annual rainfall ranges from 1900 to 3000 mm, concentrated during the wet season (October–April). This setting is highly sensitive to intense tropical rainfall, as high seasonal precipitation promotes soil saturation, elevates pore-water pressure, and reduces effective shear strength. These processes collectively increase slope instability and predispose the area to rainfall-induced slope failures.

Figure 1: Study area. ➤ p. 183

Figure 2: a) Example of a shallow landslide causing road collapse along the Jember-Banyuwangi route in the Gumitir Mountain Area, and b) damage to road infrastructure, including guardrails and exposed utility cables, following a landslide event.


Java
East Java
Banyuwangi
113°54'10"E
113°55'20"E
113°56'30"E
113°57'40"E
113°58'50"E
8°14'10"S
8°15'20"S
8°16'30"S
8°17'40"S
Legend
Testing
Training
Road
Study Area
Elevation (m)
3278
3
0
0.5
1
2
3
km
N
Scale : 1 : 50.000
Content by : Entin Hidayah
Map by : Afif Amiluddin
Source : SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah

3 Research methodologies

The research implementation is divided into three stages: (1) data preparation and processing, (2) identifying the factors that condition landslides based on landslide susceptibility mapping, and (3) determining landslide mitigation strategies.

The spatial dataset comprised 96 landslide occurrence points and 180 non-landslide points, derived from field surveys, remote sensing interpretation, and ancillary data. To evaluate the predictive performance of the models, the dataset was randomly divided into two subsets: 70% for model training and 30% for testing. This partitioning ensured that the model calibration and validation processes were conducted on mutually exclusive data, thereby minimizing the risk of over-fitting and enabling an unbiased assessment of predictive capability.

Two statistical and machine learning approaches, LR and RF, were applied to establish the relationship between landslide occurrences and conditioning factors.

3.1 Data preparation and processing of landslide conditioning factors

Geospatial data were obtained from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) and Landsat 8 imagery. Six factors, which have been used in previous landslide susceptibility studies, were analyzed: elevation, aspect, slope, flow accumulation (FA), topographic wetness index (TWI), and normalized difference vegetation index (NDVI) (Liu, S. et al. 2021; Feng et al. 2022; Kohno and Higuchi 2023; Okoli et al. 2023). The data sources and processing methods are listed in Table 1.

Slope is a measure of the angle of the steepness of a land surface or slope. Aspect is the prevailing slope direction of a terrain surface, indicating how much an area is exposed to environmental factors like sunlight, wind, and rainfall, which subsequently influences the rate of weathering and the soil's moisture levels (Getachew and Meten 2021). Elevation often appears as a key predictor in landslide prediction models, although it may only serve as a proxy for other, more meaningful variables such as slope curvature or aspect (Hussain et al. 2022). TWI is one of the factors frequently used in landslide susceptibility studies (Dahim et al. 2023; Fan et al. 2023). TWI measures the potential for water accumulation in an area based on topography, which can affect slope stability and trigger landslides. TWI represents the spatial wetness level of a basin, expressed by equation (1) (Beven and Kirkby 1979):

$$TWI = ln\left(\frac{\partial}{tan\beta}\right) \quad \text{Equation (1),}$$

where ∂ represents the area drained per unit length of contour at a given point, and tan β is the slope angle at the point.

The FA represents the total area of land contributing surface runoff to a specific location, where water tends to accumulate. It is computed by summing the number of upstream cells that drain into a particular downslope cell from a given reference point. The FA is obtained for each cell in the DEM grid. The NDVI is a factor that describes the vegetation density in a certain area. Bare land without vegetation is more susceptible to landslides than forests or grasslands with higher NDVI (Acharya and Lee 2019). The NDVI value represents vegetation cover from low to high, with a value range from −1 to 1, respectively. The NDVI is formulated as in equation (2):

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \quad \text{Equation (2),}$$

where NIR represents the near-infrared band, and RED represents the reflection values of the red band.

The values for each class of landslide conditioning factors, processed from various satellite image data sources, can be seen in Table 1 and Figure 3.

Figure 3: Influencing factors for landslide susceptibility mapping: a) elevation, b) aspect, c) TWI, d) FA, e) slope, f) NDVI. ➤ p. 185

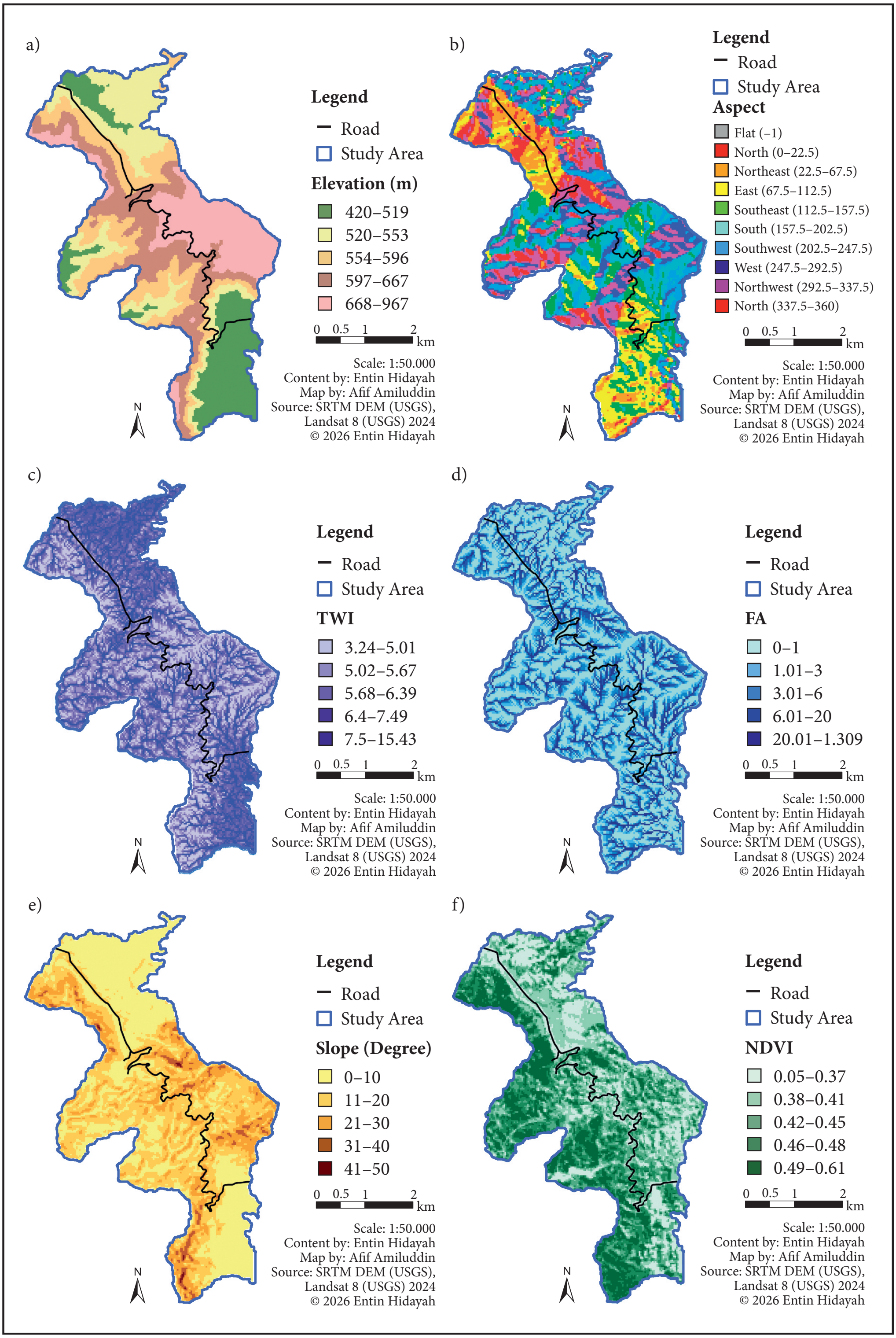

a)
Legend
Road
Study Area
Elevation (m)
420–519
520–553
554–596
597–667
668–967
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N
b)
Legend
Road
Study Area
Aspect
Flat (–1)
North (0–22.5)
Northeast (22.5–67.5)
East (67.5–112.5)
Southeast (112.5–157.5)
South (157.5–202.5)
Southwest (202.5–247.5)
West (247.5–292.5)
Northwest (292.5–337.5)
North (337.5–360)
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N
c)
Legend
Road
Study Area
TWI
3.24–5.01
5.02–5.67
5.68–6.39
6.4–7.49
7.5–15.43
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N
d)
Legend
Road
Study Area
FA
0–1
1.01–3
3.01–6
6.01–20
20.01–1.309
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N
e)
Legend
Road
Study Area
Slope (Degree)
0–10
11–20
21–30
31–40
41–50
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N
f)
Legend
Road
Study Area
NDVI
0.05–0.37
0.38–0.41
0.42–0.45
0.46–0.48
0.49–0.61
0 0.5 1 2 km
Scale: 1:50.000
Content by: Entin Hidayah
Map by: Afif Amiluddin
Source: SRTM DEM (USGS),
Landsat 8 (USGS) 2024
© 2026 Entin Hidayah
N

Table 1: Classification, source, classification method, and references for the factors.

Factor	Classes	Source (provided by)
Elevation (m)	(i) 420.00–518.67, (ii) 518.68–553.00, (iii) 553.01–595.90, (iv) 595.91–666.69, (v) 666.70–967.00	SRTM DEM: 30m × 30m United States Geological Survey (USGS) via EarthExplorer (https://earthexplorer.usgs.gov/)
Aspect (°)	Flat: -1, North: 0.0–22.5, Northeast: 22.5–67.5, East: 67.5–112.5, Southeast: 112.5–157.5, South: 157.5–202.5, Southwest: 202.5–247.5, West: 247.5–292.5, Northwest: 292.5–337.5, North: 337.5–360	Derived from DEM: 30m × 30m
Topographic wetness index (TWI)	(i) 3.24–5.01, (ii) 5.02–5.67, (iii) 5.68–6.39, (iv) 6.40–7.49, (v) 7.5–15.43	Derived from DEM: 30m × 30m equation (2)
Flow accumulation (FA)	(i) 0.0–1.00, (ii) 1.01–3.00, (iii) 3.01–6.00, (iv) 6.01–20.00, (v) 20.01–1,309	Derived from DEM: 30m × 30m
Slope (°)	(i) 0–10, (ii) 11–20, (iii) 21–30, (iv) 31–40, (v) 41–50	Derived from DEM: 30m × 30m
Normalized difference vegetation index (NDVI)	(i) 0.05 - 0.37, (ii) 0.38 - 0.41, (iii) 0.42 - 0.45, (iv) 0.46 - 0.48, (v) 0.49 - 0.61	Landsat 8 (OLITIRS) United States Geological Survey (USGS) via EarthExplorer (https://earthexplorer.usgs.gov/)

3.2 Multicollinearity testing and feature selection in landslide susceptibility modeling

Analyzing the factors causing landslides is important in susceptibility studies, as each factor may have a different influence on prediction results (Bravo-López et al. 2023; Huang, F. et al. 2023; Saha et al. 2023). To ensure model accuracy, landslide-causing factors need to be tested for multicollinearity and Spearman's rank correlation to examine relationships between variables.

In regression models, independent variables should not have perfect linear relationships with each other. High correlation among variables can cause problems in modeling and result interpretation. Therefore, multicollinearity testing is necessary, typically by calculating the Variance Inflation Factor (VIF) (Liu, R. et al. 2022; Saha et al. 2023).

Multicollinearity diagnosis is essential before modeling to ensure that no susceptibility factors are influencing each other. Potential multicollinearity in the dataset can be identified if VIF values exceed 5 or tolerance values fall below 0.1 (Huang, F. et al. 2023). If two indicators in the dataset fall within the expected range, it suggests that there is no significant multicollinearity among the landslide susceptibility factors.

Spearman correlation analysis evaluates relationships among independent variables to identify the most relevant factors, reducing uncertainty and improving model performance. Factors with high correlation coefficients are considered redundant and should be removed (Bravo-López et al. 2023). A Spearman correlation coefficient of 0.70 or lower is typically used as a cut-off point for identifying suitable features (Gu et al. 2026).

3.3 Logistic regression model

The LR model is a form of linear regression normalized by the sigmoid function (logistic function). This approach is used to predict landslide occurrence by assuming a binary response variable, i.e., occurrence (landslide occurred) and non-occurrence (no landslide), based on categorical and continuous variables (Liu, Y. et al. 2022). A value of 1 indicates a location where a landslide has taken place, while a value of 0 represents an area where no landslide occurred. The landslide triggering parameters and their relationships are calculated using the logistic equation as in equation (3):

$$P = \frac{1}{1+e^{-z}} \qquad \text{Equation (3),}$$

where P denotes the estimated likelihood of a landslide, ranging from 0 to 1 and forming an S-shaped curve, while z represents the combined effect of contributing factors like slope, as calculated using equation (4) below. Here, e is Euler's constant (approximately 2.71828), the base of the natural logarithm.

$$Z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \quad \text{Equation (4),}$$

where bi (i = 0, 1, 2, …, n) represents the LR coefficients, and xi (i = 0, 1, 2, …, n) represents the conditioning factors, with n being the number of input variables (Rong et al. 2023).

3.4 Random forest

RF has been utilized to solve complex problems in data science due to its high accuracy and interpretability. It operates effectively without the need for feature scaling or encoding categorical variables, and it typically involves minimal parameter adjustment (Sarkar et al. 2022). The RF algorithm determines its final prediction based on majority voting, aiming to minimize prediction error. It comprises numerous decision trees, each incorporating two layers of randomness. First, every tree is trained using a randomly drawn subset of the original dataset. Second, at each decision node, a random selection of features is evaluated to find the optimal split.

The overall workflow of RF involves: (1) randomly sampling subsets from the dataset; (2) constructing individual decision trees from each sample and generating predictions; and (3) aggregating the predictions, using the most common output for classification tasks and the average (mean) for regression tasks.

3.5 Model performance

The model's performance is evaluated by comparing the results of landslide susceptibility models using the area under the curve (AUC) value. AUC quantifies the area beneath the receiver operating characteristic (ROC) curve, which is a plot showing the model's performance by comparing the true positive rate (TPR) or sensitivity against the False Positive Rate (FPR) or 1-Specificity across different thresholds. The ROC curve helps assess the model's ability to differentiate between positive and negative classes, with a curve closer to the top-left corner indicating better performance. AUC values range from 0.5 to 1, where a higher AUC indicates superior model performance. An AUC of 0.5 suggests random guessing, while values closer to 1 reflect excellent predictive capability. In this study, the results of two models were validated using 48 landslides, and the accuracy of the model predictions was further assessed by calculating the proportion of landslide pixels in each vulnerability zone.

4 Results

4.1 Factors that most significantly influence landslide susceptibility in the Gumitir Mountain Area

The results of the multicollinearity test for all factors (Table 2) show that all VIF values are less than 5, and tolerance values exceed 0.1, indicating that all variables are independent and there is no potential multicollinearity among them. Therefore, the six factors, i.e., slope, aspect, elevation, TWI, FA, and NDVI, can be used for feature selection.

Otherwise, the conditioning factors need to be tested to meet the model analysis requirements for factor dependencies. Based on Table 3, Spearman's correlation test indicates that all factors have values below 0.7, suggesting that there is no interdependence among the factors. The six selected environmental factors, elevation, slope, aspect, FA, NDVI, and TWI, have different levels of influence on landslide contribution.

Table 2: Multicollinearity analysis.

Term	Elevation	Slope	Aspect	FA	NDVI	TWI
VIF	1.402	1.446	1.193	1.279	1.035	1.573
Tolerance	0.713	0.692	0.838	0.782	0.966	0.636

Table 3: Spearman correlation.

Factor	Elevation	Slope	Aspect	FA	NDVI	TWI
Elevation	1.000	0.521	0.177	−0.047	0.057	−0.376
Slope		1.000	−0.004	0.045	0.145	−0.488
Aspect			1.000	−0.018	−0.146	−0.047
FA				1.000	0.001	0.429
NDVI					1.000	−0.102
TWI						1.000

Therefore, these factors can be used in the classification process using RF, considering that the response variable is categorical.

4.2 Comparing the LR and RF applications for landslide susceptibility

Slope was identified as the primary factor affecting landslide susceptibility in the RF model, with an average merit value of 0.75, followed by FA, 0.7, and NDVI (0.64), elevation (0.6), and TWI (0.57). Aspect (0.54) exhibited comparatively lower contributions (Figure 4). These results underscore the critical role of slope steepness and hydrological conditions (FA) in driving landslide risks, while highlighting the secondary but non-negligible influence of vegetation cover and soil moisture dynamics.

The LR model identified specific thresholds for landslide triggers (Figure 5). The analysis was based on the odds ratios of several key factors influencing landslide susceptibility, including elevation, slope, aspect, FA, NDVI, and TWI. A larger odds ratio signifies a greater probability of a landslide occurring as the value of the factor increases. Areas with elevations between 662.04 m and 702.27 m show the highest odds ratio (8.1), indicating that these areas have the greatest chance of experiencing landslides compared to other elevation classes. Conversely, elevations below or above this range have lower odds ratios.

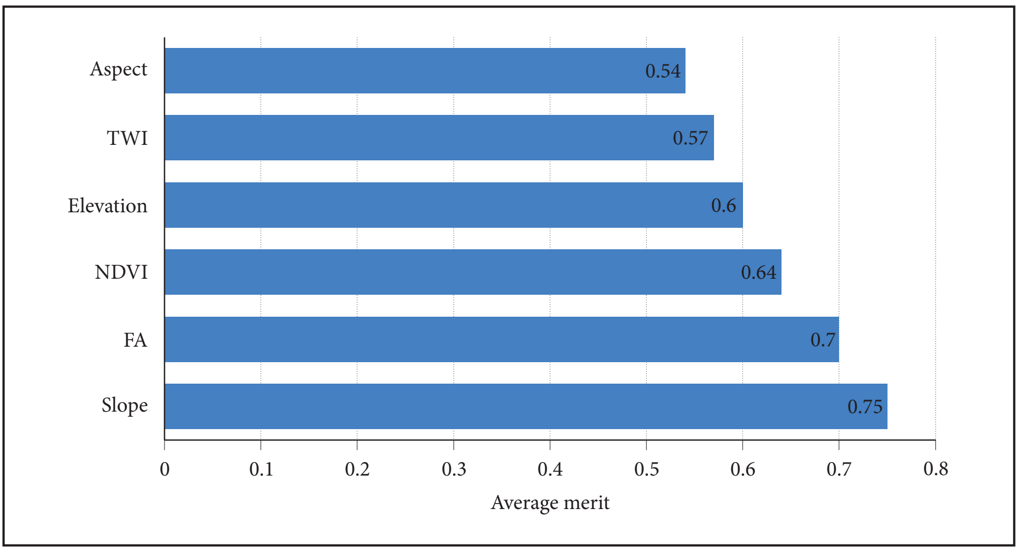
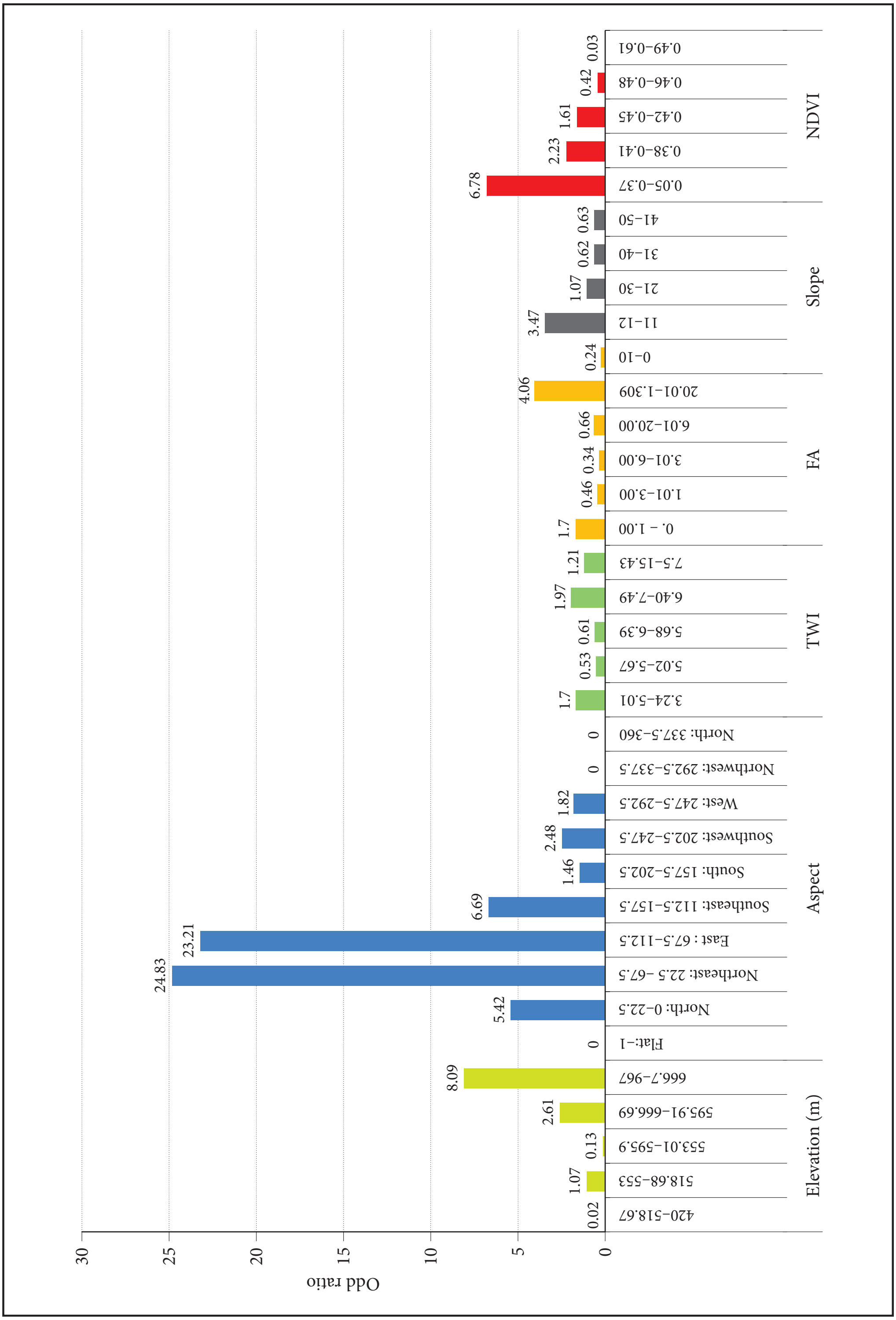


Figure 4: Average merit of the RF model.

Figure 5: Odds ratio of the logistic regression method. ➤ p. 189


Odd ratio
0
5
10
15
20
25
30
Elevation (m)
420–518.67
518.68–553
553.01–595.9
595.91–666.69
666.7–967
0.02
1.07
0.13
2.61
8.09
Aspect
Flat:-1
North: 0–22.5
Northeast: 22.5 –67.5
East : 67.5–112.5
Southeast: 112.5–157.5
South: 157.5–202.5
Southwest: 202.5–247.5
West: 247.5–292.5
Northwest: 292.5–337.5
North: 337.5–360
0
5.42
24.83
23.21
6.69
1.46
2.48
1.82
0
0
TWI
3.24–5.01
5.02–5.67
5.68–6.39
6.40–7.49
7.5–15.43
1.7
0.53
0.61
1.97
1.21
FA
0. – 1.00
1.01–3.00
3.01–6.00
6.01–20.00
20.01–1.309
1.7
0.46
0.34
0.66
4.06
Slope
0–10
11–12
21–30
31–40
41–50
0.24
3.47
1.07
0.62
0.63
NDVI
0.05–0.37
0.38–0.41
0.42–0.45
0.46–0.48
0.49–0.61
6.78
2.23
1.61
0.42
0.03

Slopes with a gradient of 10–20° show a high odds ratio of 3.5, followed by slopes with gradients of 0–10° (1.1) and 20–30° (1.0). This suggests that slopes with moderate gradients are more susceptible to landslides compared to very steep or very gentle slopes.

Aspects with orientations of 67.5–112.4° show a very high odds ratio of 24.8, and aspects of 112.5–157.4° have an odds ratio of 23.2. This indicates that slopes facing certain directions have a significantly higher landslide risk.

The highest FA values, between 2.92 and 11.3, correspond to an odds ratio of 4.1, suggesting that areas with elevated water FA are more susceptible to landslides. Areas with TWI values between 6.40 and 7.49 exhibit the highest odds ratio (1.97), indicating that locations with greater moisture accumulation and soil saturation tend to be more susceptible to landslides due to reduced soil shear strength and increased pore-water pressure.

While both models confirmed slope and hydrology as key drivers, RF provided granular insights into variable importance, whereas LR quantified risk thresholds for specific parameter ranges. RF prioritized slope stability (merit = 0.75), aligning with LR's identification of moderate slopes (10–20°) as high-risk zones. LR revealed extreme susceptibility of northeast-facing slopes (odds ratio > 20), a nuance not explicitly captured by RF's lower aspect merit (0.54). Both models emphasized the protective role of vegetation (RF: NDVI merit = 0.64; LR: low NDVI odds ratio = 6.8).

These findings highlight the complementary strengths of RF (variable ranking) and LR (threshold-specific risk quantification) in guiding targeted interventions. For example, stabilizing moderate slopes (10–20°) and restoring vegetation in low-NDVI areas could reduce risks by addressing both RF- and LR-identified vulnerabilities.

4.3 Landslide susceptibility map

The performance of the LR and RF models was evaluated using the ROC curve and AUC for the testing dataset (Figure 6). The LR model shows accuracy with an AUC value of 0.931, while the RF model has an AUC value of 0.865. This indicates that spatial landslide susceptibility analysis using both RF and LR models provides outstanding accuracy and can be relied upon for predicting landslides.

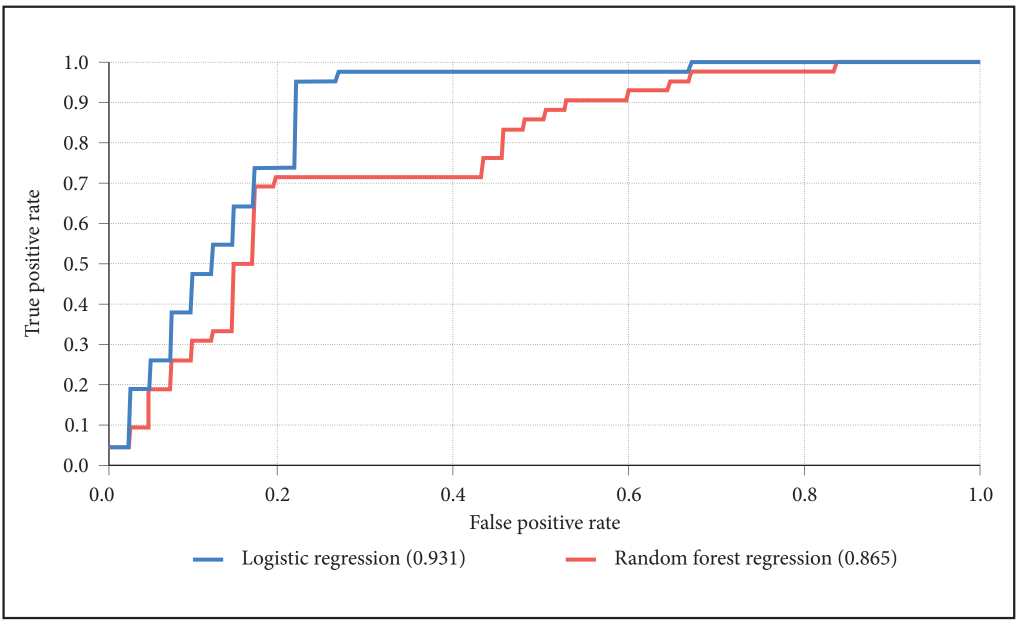


Figure 6: ROC curves illustrating the performance of the LR and RF models.

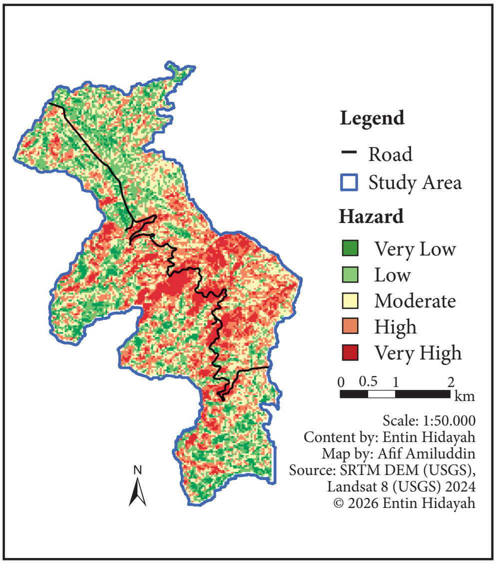


Figure 7: Landslide susceptibility mapping using RF.

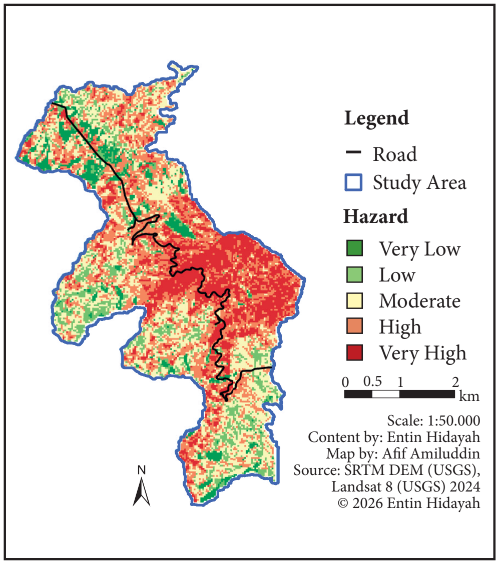


Figure 8: Landslide susceptibility mapping using LR.

Landslides with very high indices are primarily concentrated in the central area of the study region and along the road. The continuous landslide susceptibility index values generated by both models were then reclassified into five hazard classes (Very Low, Low, Moderate, High, and Very High) using the Natural Breaks (Jenks) classification method. This method identifies natural groupings inherent in the data to create an optimal classification. The final susceptibility maps for the RF and LR models are presented in Figure 7 and Figure 8, respectively. In the Gumitir mountain area, landslides are distributed on both sides of the road.

4.4 Proposed mitigation strategies

Based on the RF and LR models, which highlight slope, FA, NDVI, TWI, and aspect as the most critical factors, three integrated strategies are proposed to mitigate shallow landslides in the Gumitir Mountain Area.

The first strategy is slope stabilization and soil management. Terracing on moderate slopes (10–20°) reduces runoff and shear stress, while retaining structures in high FA areas (>2.9) counter hydrostatic pressure. These measures enhance slope resistance and mitigate landslide triggers.

The second strategy emphasizes hydrological control through drainage optimization by constructing swales or subsurface drains in zones with high TWI (>5.78) to divert excess water and balance soil moisture, combined with regular maintenance of natural waterways in high FA areas to avoid blockages during extreme rainfall events.

The third strategy involves vegetation and land use planning, prioritizing reforestation in low-NDVI zones (<0.2) using deep-rooted species such as vetiver grass and bamboo, as well as fast-growing shrubs like Calliandra to improve soil cohesion and ground cover. In addition, aspect-based zoning should restrict development on northeast–southeast-facing slopes (67.5–157.4°), which are highly susceptible due to solar radiation-induced weathering and moisture fluctuations, while maintaining these areas as buffer green belts.

Together, these slope management, hydrological interventions, and ecological restoration measures provide a comprehensive framework for reducing shallow landslide risks in the Gumitir Mountain Area while maintaining ecosystem stability and community resilience.

5 Discussion

The RF-LR synergy provides actionable insights, while RF prioritizes slope and FA, and LR identifies critical thresholds (e.g., elevation bands, aspect angles). In practice, landslide mitigation techniques are often applied partially, focusing on locations that are frequently impacted by landslides. This approach offers only a temporary fix and overlooks root causes, so landslide management needs a holistic understanding to avoid overlooked factors.

Slope angle has been proven to be a dominant predictor of shallow landslides in the RF model (merit value = 0.75) and in the LR model (odds ratio = 3.5 for slopes between 10–20°). Steeper slopes tend to have thinner soil, which increases the role of cohesion and boundary stress, thereby reducing the frequency of landslides despite the additional gravitational pull (Dehnavi et al. 2015; Prancevic et al. 2020). Steeper slopes increase landslide potential because the shear stress on steep slopes tends to rise, as also explained by Ghasemian et al. (2022). Generally, landslides often occur in areas with steep slopes greater than 25° (Ghasemian et al. 2022; Zárate et al. 2023). However, in this case, the slope angle in the range of 10–20° has proven to have a significant impact on shallow landslide occurrences during the rainy season. Landslide characteristics on lower slope gradients are generally triggered by high-intensity rainfall, which increases pore pressure in the soil and reduces soil particle cohesion, eventually triggering soil movement (Liu, Y. et al. 2021). Therefore, slopes that are not very steep are often considered relatively stable, and landslide risks on these slopes are frequently overlooked.

The FA also shows a significant influence on both methods. Water FA plays a crucial role in triggering landslides, especially in areas with high FA values. Rainfall increases soil moisture content and reduces matric suction, leading to a decline in slope stability and triggering landslides. Preferential and matrix flow in unsaturated conditions play important roles in this process (Petrella et al. 2023; Sun et al. 2023). Hydrogeological processes such as rainfall infiltration and groundwater circulation affect slope stability. Water flowing through large pores and cracks can accelerate the landslide process (Bishop et al. 2021; Petrella et al. 2023). In vulnerable areas, especially around water channels, rainwater tends to accumulate and increase hydrostatic pressure, reducing the shear strength, which in turn weakens the soil structure. This condition becomes even more critical during the rainy season when high rainfall causes a drastic increase in FA, magnifying the likelihood of landslides. Therefore, concentrated rainwater can quickly increase pore water pressure, which weakens the soil shear strength, especially in areas with poor drainage or irregular water channels.

The NDVI is an important indicator in determining landslide risk potential, with low NDVI indicating areas with minimal or no vegetation. The NDVI and rainfall are dynamic factors that influence landslides, with higher rainfall and lower NDVI increasing landslide threats (Huang, W. et al. 2023) because vegetation functions as a soil binder through root systems that strengthen slope stability, but only for shallow depth (< 1m) (Ni et al. 2018). Without adequate vegetation, steep slopes and areas with high FA become more vulnerable to soil failure. Therefore, monitoring NDVI in critical locations before the rainy season is crucial to mitigating landslide risks. Areas with low NDVI combined with high slope angles and large FA require special attention for preventive measures, such as planting stabilizing vegetation or soil conservation techniques to enhance slope stability and reduce disaster potential.

Based on TWI, variations in slope stability significantly influence landslide hazard and risk, as indicated by high and low odds ratios. High soil moisture can reduce soil cohesion and weaken slope shear strength, making it more vulnerable to failure, especially during heavy rains that increase soil moisture content. High TWI values are more prone to landslides (Dahim et al. 2023; Zárate et al. 2023). On the other hand, low TWI is generally associated with less moisture accumulation; in some contexts, very low TWI values can indicate very dry soil conditions and potentially have cracks or moisture deficiencies that can lead to soil failure when rain occurs. Low TWI has a strong influence on shallow landslides, especially in areas with high slope gradients and high drainage density (Gullà et al. 2021). Therefore, although high TWI consistently indicates increased landslide risk, low TWI should also be considered a risk factor, particularly in locations with steep slope gradients.

Aspect shows a lower influence compared to other factors, but remains important. Slope orientation can affect the distribution of sunlight and soil moisture, which in turn can influence slope stability. Aspect or slope orientation significantly influences landslide susceptibility, especially through slope hydrology mechanisms and rainfall patterns (Guo and Ma 2023). However, an aspect's influence cannot be evaluated in isolation and must be considered alongside other factors such as elevation, slope gradient, and geological conditions (Cellek 2021).

The connections between these factors are simultaneous and interdependent in triggering landslides. For example, slope and FA may work together to increase risk, where steep slopes with accumulated water flow are more likely to experience landslides. Similarly, elevation and aspect can affect sunlight exposure and soil moisture levels, impacting vegetation conditions measured by NDVI.

The RF model can capture the complexity of these interactions by assigning different weights to each factor, while LR focuses more on the individual contribution of factors in specific classes. Using both methods simultaneously provide a more holistic understanding of landslide risk, allowing for more accurate and effective mitigation. These results highlight the model's superiority in providing deeper insights into the environmental dynamics underlying landslide risk, ultimately aiding in more effective risk mitigation decision-making. In this study, the RF model demonstrated better performance compared to LR. Similar findings were reported by Wang et al. (2020), where RF also outperformed other commonly used methods such as Frequency Ratio, Logistic Model Tree (LMT), and Classification and Regression Tree (CART) in landslide susceptibility mapping. This addition ensures consistency and provides a broader perspective on the applicability of RF in comparison with other methods.

6 Conclusion

This study demonstrates that the random forest (RF) model outperforms logistic regression (LR) in landslide susceptibility mapping by effectively capturing complex interactions among environmental factors. Slope gradient, flow accumulation (FA), vegetation density (normalized difference vegetation index, NDVI), and topographic wetness index (TWI) emerged as dominant drivers of landslide risk. Slope, with the highest merit value (RF) and elevated odds ratio (LR), underscores the vulnerability of moderate slopes (10–20°) to pore-pressure dynamics during heavy rainfall. The FA and NDVI further highlight the roles of concentrated water flow and vegetation deficits in slope destabilization, while TWI reveals that both extreme soil saturation and aridity threaten stability.

To address these risks, three integrated mitigation strategies are proposed for slope stabilization and soil management, including the implementation of terracing on moderate slopes, the planting of deep-rooted vegetation in low-NDVI areas, and the installation of geosynthetic-reinforced walls. Hydrological control measures are also recommended through the optimization of the drainage system by maintaining natural waterways to prevent blockages and overflows during rainfall events. In addition, vegetation restoration in low-NDVI areas and aspect-specific land-use zoning is proposed to restrict inappropriate development in landslide-prone areas.

Future research should focus on the application of advanced deep learning algorithms and high-resolution GIS-based modeling to better predict nonlinear interactions among environmental factors such as NDVI-aspect feedbacks and improve adaptation to climate-induced rainfall variability. Integrating socioeconomic factors including land-use practices, into predictive models could further enhance their applicability and decision-making value.

ACKNOWLEDGEMENTS: We would like to extend our sincere thanks to the Directorate General of Higher Education, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, for providing financial support for this research through the DRTPM grant in 2024. We also express our gratitude to the Institute for Research and Community Service, University of Jember, and the Water Resource Management and Engineering Laboratory for their assistance and resources in facilitating this study.

RESEARCH DATA: For information on the availability of research data related to the study, please visit the article webpage: https://doi.org/10.3986/AGS.14449.

7 References

Abdulwahid, W. M., Pradhan, B. 2017: Landslide vulnerability and risk assessment for multi-hazard scenarios using airborne laser scanning data (LiDAR). *Landslides* 14-3. https://doi.org/10.1007/s10346-016-0744-0

Acharya, T. D., Lee, D. H. 2019: Landslide susceptibility mapping using relative frequency and predictor rate along Araniko highway. *KSCE Journal of Civil Engineering* 23-2. https://doi.org/10.1007/s12205-018-0156-x

Achour, Y., Pourghasemi, H. R. 2020: How do machine learning techniques help in increasing accuracy of landslide susceptibility maps?. *Geoscience Frontiers* 11-3. https://doi.org/10.1016/j.gsf.2019.10.001

Akturk, E., Guney, K., Aydin, M., Yildiz, F., Kaya, O. 2024: Land cover dynamics and environmental implications in Türkiye's riparian zones: A comprehensive seven-year study. *Forest* 15-7. https://doi.org/10.3390/f15071177

Alcántara-Ayala, I. 2025: Landslides in a changing world. *Landslides* 2022. https://doi.org/10.1007/s10346-024-02451-1

Beven, K. J., Kirkby, M. J. 1979: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin* 24-1. https://doi.org/10.1080/02626667909491834

Bishop, J. L., Yeşilbaş, M., Hinman, N. W., Burton, Z. F. M., Englert, P. A. J., Toner, J. D., McEwen, A. S. et al. 2021: Martian subsurface cryosalt expansion and collapse as trigger for landslides. *Science Advances* 7-6. https://doi.org/10.1126/sciadv.abe4459

Bravo-López, E., Del Castillo, T. F., Sellers, C., Delgado-García, J. 2023: Analysis of conditioning factors in Cuenca, Ecuador, for landslide susceptibility maps generation employing machine learning methods. *Land* 12-6. https://doi.org/10.3390/land12061135

Cellek, S. 2021: The effect of aspect on landslide and its relationship with other parameters. In: Landslides. IntechOpen. I. https://doi.org/10.5772/intechopen.99389

Dahim, M., Alqadhi, S., Mallick, J. 2023: Enhancing landslide management with hyper-tuned machine learning and deep learning models: Predicting susceptibility and analyzing sensitivity and uncertainty. *Frontiers in Ecology and Evolution* 11. https://doi.org/10.3389/fevo.2023.1108924

Dehnavi, A., Aghdam, I. N., Pradhan, B., Varzandeh, M., Morshed, H. 2015: A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena* 135. https://doi.org/10.1016/j.catena.2015.07.020

Fan, H., Lu, Y., Shao, S., Li, L., Wang, Y., Lu, M., Li, J. et al. 2023: Evaluation and analysis of statistical and coupling models for highway landslide susceptibility. *Geomatics, Natural Hazards and Risk* 14-1. https://doi.org/10.1080/19475705.2023.2167612

Feng, L., Guo, M., Wang, W., Chen, Y., Shi, Q., Guo, W., Lou, Y. et al. 2022: Comparative analysis of machine learning methods and a physical model for shallow landslide risk modeling. *Sustainability* 15-6. https://doi.org/10.3390/su15010006

Getachew, N., Meten, M. 2021: Weights of evidence modeling for landslide susceptibility mapping of Kabi-Gebro locality, Gundomeskel area, Central Ethiopia. *Geoenvironmental Disasters* 8-1. https://doi.org/10.1186/s40677-021-00177-z

Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Kress, V., Renoud, S. et al. 2022: A robust deep-learning model for landslide susceptibility mapping: A case study of Kurdistan Province, Iran. *Sensors* 22. https://doi.org/10.3390/s22041573

Gullà, G., Conforti, M., Borrelli, L. 2021: A refinement analysis of the shallow landslides susceptibility at regional scale supported by GIS-aided geo-database. *Geomatics, Natural Hazards and Risk* 12-1. https://doi.org/10.1080/19475705.2021.1967204

Guo, Y., Ma, C. 2023: Elucidating the role of soil hydraulic properties on aspect-dependent landslide initiation. *Hydrology and Earth System Sciences* 27-8. https://doi.org/10.5194/hess-27-1667-2023

Gu, R., Ding, H., Yu, H., Wang, Q., Kong, L., Fan, J. 2026: Incorporating spatial heterogeneity into landslide susceptibility assessment: a GWRF–SHAP new method. *International Journal of Digital Earth* 19-1. https://doi.org/10.1080/17538947.2026.2639808

Huang, F., Teng, Z., Guo, Z., Catani, F., Huang, J. 2023: Uncertainties of landslide susceptibility prediction: Influences of different spatial resolutions, machine learning models, and proportions of training and testing dataset. *Rock Mechanics Bulletin* 2-1. https://doi.org/10.1016/j.rockmb.2023.100028

Huang, W., Ding, M., Li, Z., Yu, J., Ge, D., Liu, Q., Yang, J. 2023: Landslide susceptibility mapping and dynamic response along the Sichuan-Tibet transportation corridor using deep learning algorithms. *Catena* 222. https://doi.org/10.1016/j.catena.2022.106866

Hussain, M. A., Chen, Z., Shah, S. U., Shoaib, M., Ali, N., Ma, C. 2022: Landslide susceptibility mapping using machine learning. *Civil Engineering Journal* 8-2. https://doi.org/10.28991/CEJ-2022-08-02-02

Kohno, M., Higuchi, Y. 2023: Landslide susceptibility assessment in the Japanese Archipelago based on a landslide distribution map. *ISPRS International Journal of Geo-Information* 12-2. https://doi.org/10.3390/ijgi12020037

Le, X. H., Choi, C., Eu, S., Yeon, M., Lee, G. 2024: Quantitative evaluation of uncertainty and interpretability in machine learning-based landslide susceptibility mapping through feature selection and explainable AI. *Frontiers in Environmental Science* 12. https://doi.org/10.3389/fenvs.2024.1424988

Liu, R., Yang, X., Xu, C., Wei, L., Zeng, X. 2022: Comparative study of convolutional neural network and conventional machine learning methods for landslide susceptibility mapping. *Remote Sensing* 14-2. https://doi.org/10.3390/rs14020321

Liu, S., Yin, K., Zhou, C., Gui, L., Liang, X., Lin, W., Zhao, B. 2021: Susceptibility assessment for landslide initiated along power transmission lines. *Remote Sensing* 13-24. https://doi.org/10.3390/rs13245068

Liu, Y., Deng, Z., Wang, X. 2021: The effects of rainfall, soil type and slope on the processes and mechanisms of rainfall-induced shallow landslides. *Applied Sciences* 11-24. https://doi.org/10.3390/app112411652

Liu, Y., Zhao, L., Bao, A., Li, J., Yan, X. 2022: Chinese high resolution satellite data and GIS-based assessment of landslide susceptibility along highway G30 in Guozigou valley using logistic regression and MaxEnt model. *Remote Sensing* 14-15. https://doi.org/10.3390/rs14153620

Munawar, H. S., Hammad, A. W. A., Waller, S. T. 2022: Remote sensing methods for flood prediction: A review. *Sensors* 22-3. https://doi.org/10.3390/s22030960

Ni, J. J., Leung, A. K., Ng, C. W. W., Shao, W. 2018: Modelling hydro-mechanical reinforcements of plants to slope stability. *Computers and Geotechnics* 95. https://doi.org/10.1016/j.compgeo.2017.09.001

Okoli, J., Nahazanan, H., Nahas, F., Kalantar, B., Shafri, H. Z. M., Khuzaimah, Z. 2023: High-resolution Lidar-derived DEM for landslide susceptibility assessment using AHP and fuzzy logic in Serdang, Malaysia. *Geosciences* 13-2. https://doi.org/10.3390/geosciences13020034

Petrella, E., Raimondo, M., Chelli, A., Valentino, R., Severini, E., Diena, M., Celico, F. 2023: Processes and factors controlling the groundwater flow in a complex landslide: A case study in the northern Italy. *Hydrological Processes* 37-5. https://doi.org/10.1002/hyp.14891

Prancevic, J. P., Lamb, M. P., McArdell, B. W., Rickli, C., Kirchner, J. W. 2020: Decreasing landslide erosion on steeper slopes in soil-mantled landscapes. *Geophysical Research Letters* 47-10. https://doi.org/10.1029/2020GL087505

Rong, G., Li, K., Tong, Z., Liu, X., Zhang, J., Zhang, Y., Li, T. 2023: Population amount risk assessment of extreme precipitation-induced landslides based on integrated machine learning model and scenario simulation. *Geoscience Frontiers* 14-3. https://doi.org/10.1016/j.gsf.2023.101541

Rosly, M. H., Musa, H., Mohamad, B., Bolong, N. 2023: Relationship of rainfall intensity with slope stability. *Civil Engineering Journal* 9. https://doi.org/10.28991/CEJ-SP2023-09-06

Saha, S., Bera, B., Kumar, P., Sengupta, D., Bhattacharjee, S., Sengupta, N., Majumdar, P. et al. 2023: Geosystems and geoenvironment modelling and predicting of landslide in Western Arunachal Himalaya, India. *Geosystems and Geoenvironment* 2-2. https://doi.org/10.1016/j.geogeo.2022.100158

Park, S., Hamm, S. Y., Kim, J. 2019: Performance evaluation of the GIS-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling. *Sustainability* 11-20. https://doi.org/10.3390/su11205659

Saha, S., Gayen, A., Bayen, B. 2022: Deep learning algorithms to develop flood susceptibility map in data-scarce and ungauged river basin in India. *Stochastic Environmental Research and Risk Assessment* 36. https://doi.org/10.1007/s00477-022-02195-1

Sarah, D., Zulfahmi, Z., Hilmi, M., Putra, Z., Madiutomo, N., Gunawan, G., Sumaryadi, S. et al. 2024: Back analysis of rainfall-induced landslide in Cimanggung District of Sumedang Regency in West Java using deterministic and probabilistic analyses. Geosciences 14-12. https://doi.org/10.3390/geosciences14120347

Sarkar, S. K., Ansar, S. Bin, Ekram, K. M. M., Khan, M. H., Talukdar, S., Naikoo, M. W., Islam, A. R. T. et al. 2022: Developing robust flood susceptibility model with small numbers of parameters in highly fertile regions of Northwest Bangladesh for sustainable flood and agriculture management. *Sustainability*14-7. https://doi.org/10.3390/su14073982

Sun, Q., Zhao, L., Qin, Y., Liu, B. 2023: Experimental study on the failure of loose accumulation landslides under rainfall conditions. *Frontiers in Earth Science* 11. https://doi.org/10.3389/feart.2023.1291592

Teng, F., Mao, Y., Li, Y., Qian, S., Nanehkaran, Y. A. 2024: Comparative models of support-vector machine, multilayer perceptron, and decision tree prediction approaches for landslide susceptibility analysis. *Open Geosciences* 16-1. https://doi.org/10.1515/geo-2022-0642

Wang, Y., Sun, D., Wen, H., Zhang, H., Zhang, F. 2020: 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* 17-12. https://doi.org/10.3390/ijerph17124206