Comparison of different open-source DEMs for landslide susceptibility mapping

Abstract:

The open-source digital elevation model (DEM) is a popular dataset for regional landslide susceptibility mapping (LSM), and the choice of DEM should affect the mapping accuracy. With the advancement in remote sensing, an increasing number of global open-source DEMs have been released, with improvement in the accuracy, and the latest released data are rarely evaluated in LSM research. In this paper, DEM-based factors, including elevation, aspect, slope, plan curvature, and profile curvature, were generated from seven open-source DEMs, including Advanced Spaceborne Thermal Emission and Reflection (ASTER) V2, ASTERV3, ALOS World 3D - 30 m (AW3D30), Copernicus DEM 30m (COP) Forest and Buildings removed Copernicus DEM (FABDEM), NASADEM, and Shuttle Radar Topography Mission (SRTM). DEM-based factors were coupled with the distance to road, distance to river, land use, lithology, rain, and Normalized Difference Vegetation Index (NDVI). The significant difference between DEMs is determined by comparing the area proportion. Slope, plane curvature, and profile curvature is found to have a maximum difference of 15%~20%. K-Nearest Neighbours (KNN) and Random Forest (RF) were used for LSM with two sampling methods, namely 70% for training and 30% for testing (S1); 67% for training and 33% for testing (S2). For KNN with S1, the prediction rate is range from 0.8299 to 0.8701, with a difference of 0.0402. The difference of prediction rate is decreased to 0.0207 for S2 and 0.0258 for RF. Generally, COP has the highest prediction rate of 0.8701, 0.9254, and 0.9461 for KNN with S1, RF with S1, and S2, respectively. ASTERV2 is the worst with prediction rate of 0.8897, and 0.8996 for KNN with S2, and RF with S1, respectively. The research result provides a reference for open-source DEM selection in future LSM.

Keywords: *DEM selection, Open-source DEMs, Landslide susceptibility, Machine learning*

1. Introduction

Landslides are one of the most significant geological disasters worldwide. They pose a great threat to the safety of people's lives and property due to their strong destructive power, large area that can be affected, high energy of the sliding rock, fast movement speed, and long movement distance. Landslides caused approximately 28,139 deaths in China from 1950 to 2016 (Lin and Wang, 2018). The Loess Plateau has one of the most fragile geological environments due to its transitional characteristics (Peng et al., 2014). It is highly susceptible to landslides caused by extreme weather and human activities. 15,000 geological disasters have

occurred in the Loess Plateau, of which 85% were landslides (Zhuang et al., 2018). Many studies have shown that LSM plays an irreplaceable role in studying the distribution of landslides and predicting their occurrence (Ado et al., 2022; Broeckx et al., 2018; Chen and Chen, 2021).

DEMs are important factors in landslide susceptibility studies (Saleem et al., 2019), and there are multiple ways to generate DEMs (Okolie and Smit, 2022; Xiong et al., 2022), such as interpolation from existing contour maps, digital photogrammetry, interferometric synthetic aperture radar, and light detection and ranging, all of which have been applied to varying degrees in landslide susceptibility mapping (Dou et al., 2019; Hussain et al., 2021; Yalcin and Bulut, 2007). For the differences in landslide susceptibility mapping with multiple DEMs, Arabameri et al. (2019). They found that the accuracy of PALSAR is higher than that of ASTERV1. Rabby et al. (2020) compared the landslide susceptibility mapping performances of 30 m ASTER, 30 m SRTM, 12.5 m Advanced Land Observing Satellite, and 25 m Survey of Bangladesh (SOB) DEM in Bangladesh. They found that except for poor SOB, the other three performed similarly. Mahalingam and Olsen (2016) studied the impact of DEM sources on landslide susceptibility using ASTER, the US National Elevation Dataset, and LiDAR data. They found that LiDAR has the best performance. Brock et al. (2020) investigated the relationship between DEM quality and landslide susceptibility using TanDEM-X, ASTERV2, SRTM, and Contour DEM and concluded that TanDEM-X is the best and ASTERV2 is the worst. However, no one has ever compared the differences among multiple open-source DEMs at the most commonly used 30m resolution and given recommendations on which 1 arc open-source DEM is suitable for LSM research. Also, DEMs released after 2019 (including FABDEM, ASTERV3, AW3D30, COP and NASADEM) have not been used in LSM as of the time of this writing, and these data may improve the performance of landslide susceptibility maps. So, this study selected seven high-precision and commonly used open-source DEMs, including ASTERV2, ASTERV3, AW3D30, COP, FABDEM, NASADEM, and SRTM, to study the effects of their derived factors on landslide susceptibility and to evaluate how to choose a suitable DEM for landslide susceptibility research.

Landslide susceptibility maps help predict future landslides based on previous landslide events, identifying areas within the study region that are more likely to experience landslides and contributing to disaster prevention and early warning (Guzzetti et al., 1999; Huang et al., 2023; Li et al., 2017). There are three common types of LSM models: qualitative models, statistical models, and machine learning models. Qualitative models use expert opinions and subject knowledge to assign weights to various landslide factors to quantify landslide susceptibility. The analytic hierarchy process (AHP) is a commonly used qualitative model (Kayastha et al., 2013; Panchal and Shrivastava, 2022). Statistical models are used to reduce subjectivity in weight assignment and provide more objective and repeatable results. These models can be either bivariate or multivariate, with common methods including the weight of evidence model (Batar and Watanabe, 2021), frequency ratio model (Neupane et

al., 2023), and fuzzy logic model (Pourghasemi et al., 2012b) for bivariate analysis and logistic regression analysis (Ayalew and Yamagishi, 2005) and the artificial neural network method (Lee et al., 2004; 2003) for multivariate analysis. In recent years, more researchers have used machine learning methods for landslide susceptibility studies (Merghadi et al., 2020; Sun et al.), with commonly used algorithms including support vector machine, decision tree, and random forest (Hong et al., 2018; Peng et al., 2014; Zhou et al., 2021). Compared to traditional statistical models, machine learning may perform better in some cases due to multiple training iterations, but some machine learning models have difficulty explaining the relationship between output results and input data.

Also, many scholars have focused on the study of other factors, starting with the selection of environmental factors. Common environmental factors include several major categories, including geology, soils, hydrology, geomorphology, and land use (van Westen et al., 2008). Each major category has many derived factors, such as slope, aspect, rock type, soil type, soil moisture, geomorphic unit, land use, and roads. Different combinations of factors tend to have a certain impact on the results. Zhou et al. (2021) studied the advantages and disadvantages of different factor combinations and optimized the model based on the selected factors. After selecting the factors, the subdivision of subclasses (Yan et al., 2019), the choice of resolution (Chen et al., 2020; Yan et al., 2023), and the division of geomorphic units (Hua et al., 2021; Schlögel et al., 2018; Erener and Düzgün, 2012; Martinello et al., 2021) all have a considerable impact on the results.

In this research, two machine learning (ML)methods have been chosen to predict the landslide susceptibility simultaneously, namely KNN and RF. Combining it with the most comprehensive and up-to-date open-source DEM, a 30m resolution landslide susceptibility map was drawn in the Yulin area of the Loess Plateau, providing some suggestions for the selection of DEMs in future LSM. Open-source DEMs were chosen because the study area is large at the municipal level, and open-source DEMs are easy to obtain and are widely used in landslide susceptibility research, with high representativeness. Based on previous studies, this paper selected seven common 1 arc open-source DEMs (Table 1).

Table 1 Introduction of selected DEMs

	ASTERV2	*ASTERV3*	*AW3D30*	*COP*	*FABDEM*	*SRTM*	*NASADEM*
Scope	*83°N-83°S*	*83°N-83°S*	*82°N-82°S*	*90°S-90°N*	*60°S-80°N*	*60°N-56°S*	*60°N-56°S*
Resolution	*1 arc*	*1 arc*	*1 arc*	*1 arc*	*1 arc*	*1 arc*	*1 arc*
Precision (RMSE)	*12.6 m (Tachikawa et al., 2011)*	*8.52 m (Gesch et al., 2016)*	*7.87 m (Uuemaa et al., 2020)*	*6.73m (Li et al., 2022)*	*3-14 m (Dandabathula et al., 2022)*	*8.7 m (Mouratidis et al., 2010)*	*6.39-12.08 m (Buckley et al., 2020)*
Data collection	*2000-2009 &*	*2000-2011 &*	*2006-2011*	*2011-2015*	*2010-2015*	*2000*	*2000 & 2000-2009*

Time	*2009-2011*	*2011-2019*					
Release time	*2011*	*2019*	*2019*	*2019*	*2021*	*2002*	*2020*
Source	*NASA & METI*	*NASA & METI*	*JAXA*	*ESA*	*University of Bristol*	*NASA & NIMA*	*NASA*
Website	*https://www.gscloud.cn/*	*https://www.gscloud.cn/*	*http://www.eorc.jaxa.jp/ALOS/en/aw3d30*	*https://registry.opendata.aws/copernicus-dem/*	*https://data.bris.ac.uk/data/dataset*	*https://search.earthdata.nasa.gov/search/*	*https://search.earthdata.nasa.gov/search/*

2. Study area and datasets

2.1 Study area

Yulin is located in the northernmost part of Shaanxi Province, China, at the junction of the Loess Plateau and the Mu Us Desert, serving as a transition zone between the Loess Plateau and the Inner Mongolian Plateau. The entire study area is between 36°57'-39°35' N and 107°28'-111°15' E, with a maximum length of 309 km from east to west and a maximum width of 295 km from north to south, covering a total land area of almost 43,000 km^2(as is shown in Figure 1).

The study area is characterized by transitional zones in geology, landforms, vegetation, and climate, with significant variations within the region. Yulin is part of the middle and northern parts of the Ordos slope of the North China Platform and the Shaanbei platform of the depression. The average elevations in the southwest are 1600-1800 m, while the rest of the area has average elevations of 1000-1200 m. The highest point is located in the southwest in the Baiyu Mountains, and the lowest point is at the mouth of the Wuding River in the southeast. The northern part of this area is a sandy and flat area, covering 42% of the total land area, with the southern part being a hilly and gully area covered by loess, accounting for 58% of the area (Wei et al., 2011). Vegetation transitions from forest grassland in the southeast to dry grassland and desert grassland in the northwest (Shi et al., 2003). The climate in the study area is characterized by dry and windy springs, hot and short summers, rainy autumns, and cold and long winters (Yang et al., 2004). The study area is part of the warm-temperate and semiarid continental monsoon climate, with annual average precipitation values of 300 mm-500 mm (Sun et al., 2012). Rain in the study area is concentrated in the three months of July, August, and September, accounting for more than 60% of the annual total. Due to the combined effect of these factors, Yulin is prone to geological hazards and severe soil erosion, with frequent occurrences of landslides and collapses.

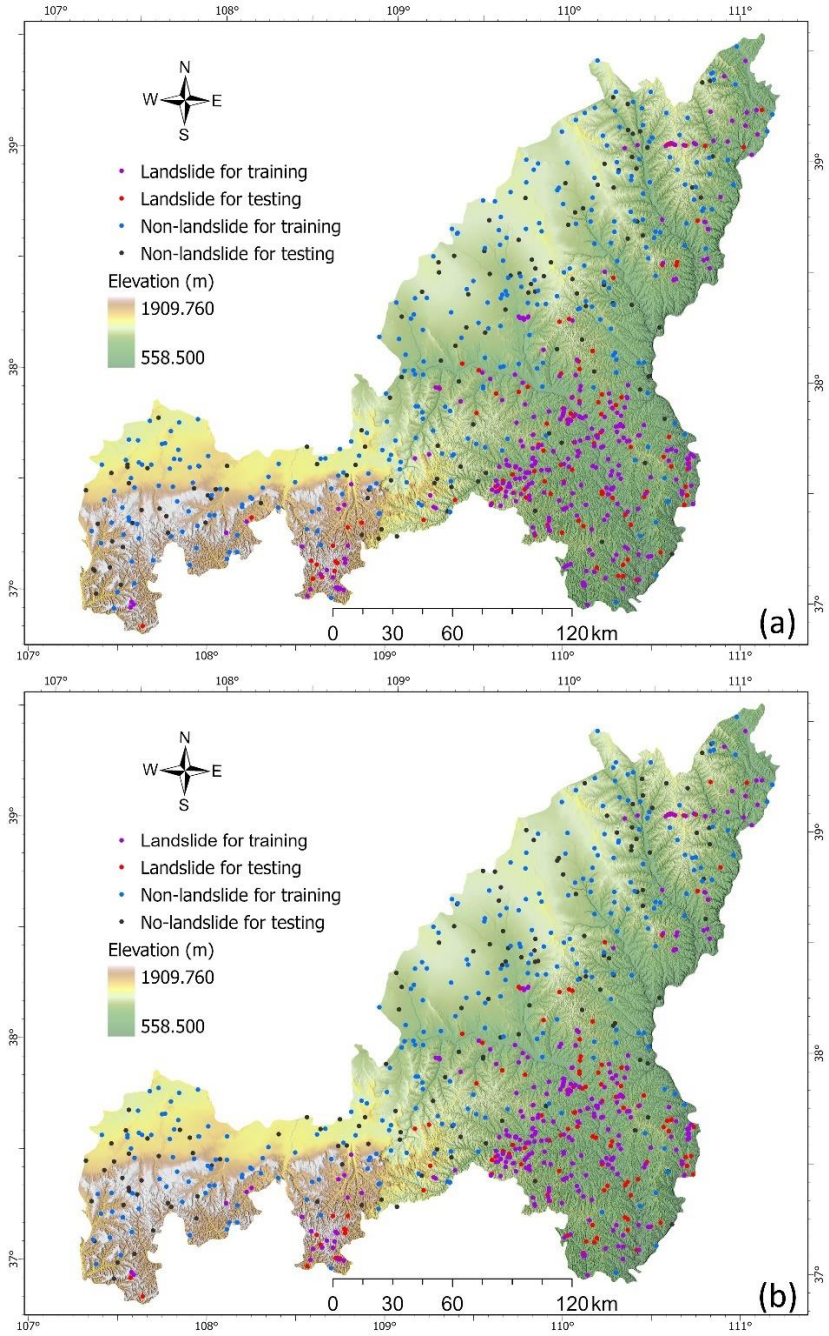

Figure 1 Study area with landslide and non-landslide points both for training and testing (a: S1, b: S2)

2.2 Datasets

The overall workflow is shown in Figure 2. The compilation of a landslide inventory map (Rohan et al., 2021) is the most basic and fundamental step in LSM. A detailed landslide inventory map enables us to fully understand the temporal and spatial distribution and variation in landslides in the study area, providing a basic data source and fundamental knowledge for subsequent research. This study compiled a landslide inventory map for the Yulin area, which includes a total of 407 landslides from 2002 to 2007. The map is based on the field investigation data of the

comprehensive evaluation and zoning of national geological disaster-prone areas conducted by the China Geological Survey Bureau in 2006. During the collection process, all landslide points are replaced by the coordinates of their center points. The minimum landslide area is 365 m^2, the maximum area is 46861 m^2, and the average area is 26532 m^2. The original data were stored in a point annotation file in WT format of MapGIS. This format is commonly used for storing geospatial data and annotations in the MapGIS software system. It was converted to SHP files in ArcGIS Pro 3.0. We randomly sampled non-landslide points from non-landslide areas. Figure 1 shows the distribution of landslides points and non-landslide points in the study area.

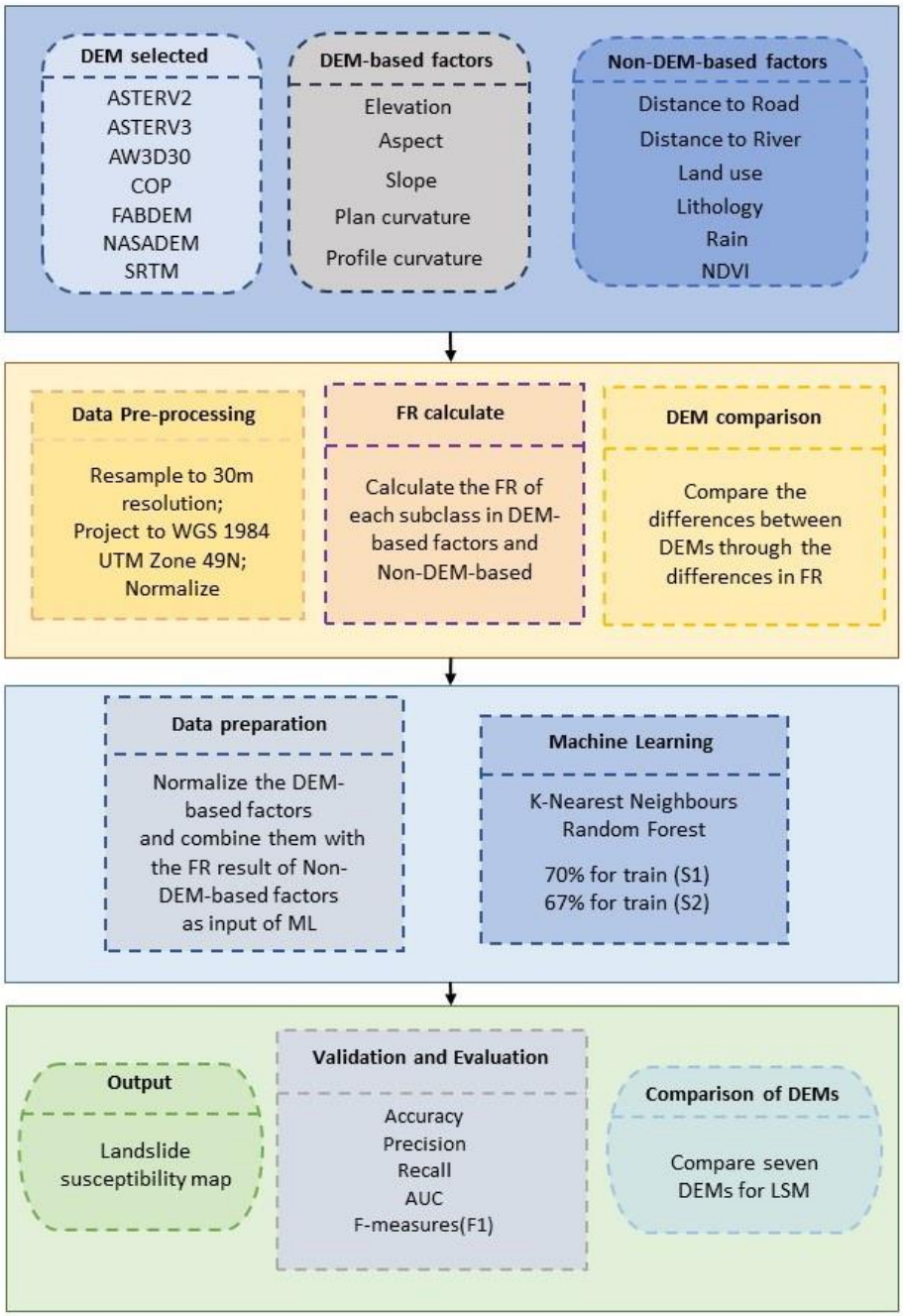

Figure 2 Flowchart of the study

Numerous studies have focused on the environmental factors affecting landslide occurrence, but there is still no consistent conclusion on these factors, except for the importance of DEM-derived factors, such as elevation, slope, aspect, profile curvature, and plan curvature (Kavzoglu et al., 2015; Xie et al., 2021). The choice of DEM affects the calculation of these factors (Warren et al., 2004), as well as the prediction of landslide susceptibility (Rabby et al., 2020). This study selected seven high-precision and commonly used open-source DEMs, including ASTERV2, ASTERV3, AW3D30, COP, FABDEM, NASADEM, and SRTM, to study the effects of their derived factors on landslide susceptibility and to evaluate how to choose a suitable DEM for landslide susceptibility research. Based on these seven DEMs, commonly used factors, such as elevation, aspect, slope, profile curvature, and plan curvature (Figure 3), were selected to map landslide susceptibility and calculated using the Surface Parameters tool in ArcGIS Pro. Due to their calculation based on DEMs, these factors are highly dependent on DEMs and greatly influenced by the source. In this paper, all factors calculated based on DEMs are referred to as DEM-based factors, while all factors that are not directly related to DEMs are referred to as non-DEM-based factors.

In the context of landslide susceptibility mapping, Yan et al. (2019) highlights that not only the choice of DEM and its derivative factors, but also the subdivision rules of subclasses can influence the accuracy of the maps. However, to mitigate the potential impact of subclassing, ML models do not employ subclassing of DEM-based factors.

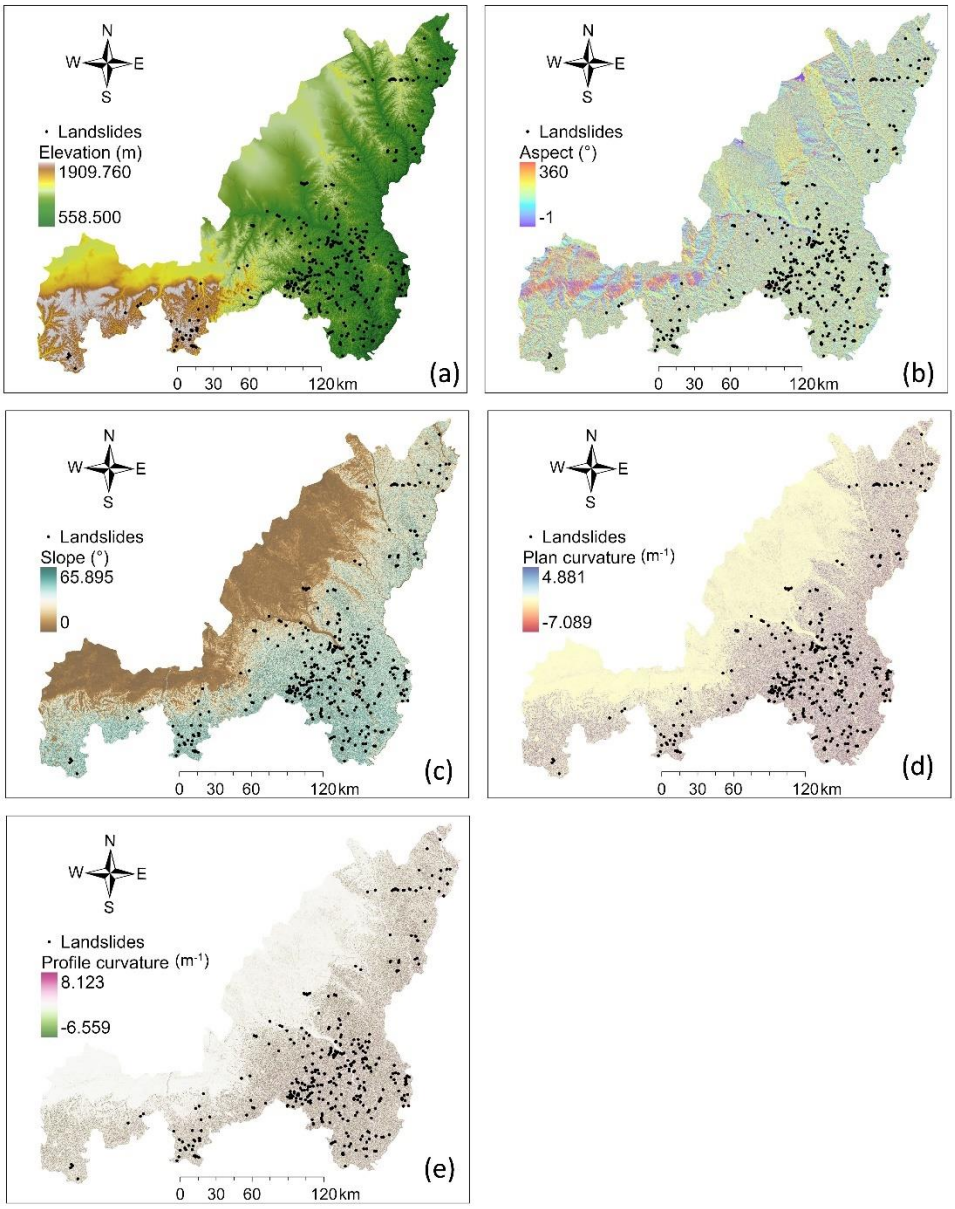

Figure 3 DEM-based factors of COP: (a) elevation, (b) aspect, (c) slope, (d) plan curvature, and (e) profile curvature

Apart from DEMs and their derived factors, several other factors have a collective impact on landslide occurrence and development. These factors include geology, climate, vegetation, and human activities. Loess, characterized by its high collapsibility, vertical joints, and susceptibility to external forces, is a fundamental material for landslides. Disturbances or sustained external forces exceeding the inherent stress can cause loess material to separate along its vertical joints, resulting in small-scale collapses or large-scale landslides (Feng et al., 2021). Additionally, concentrated heavy rainfall, particularly during summer, contributes to slope instability by eroding the slope surface (Uchida et al., 2001). The presence and abundance of vegetation also play a role in the impact of precipitation on slopes. Adequate vegetation coverage reduces surface runoff, while plant roots consolidate rock layers, enhancing soil stability and reducing landslide susceptibility (Huang et al.,

2021). Furthermore, human activities, such as engineering projects, cultivation, and deforestation (or afforestation), greatly influence the stability of the entire system, leading to an increase in the number and scale of landslides.

To incorporate the effects of geology, human activities, climate, and vegetation, five factors were selected: lithology, distance to roads, land use, distance to rivers, rain, and normalized difference vegetation index (NDVI). These factors were chosen to represent the various influences contributing to landslide occurrence and development. The distances to road and river were calculated using Euclidean Distance in ArcGIS Pro based on vector data for roads and rivers. Rainfall was generated using the cokriging interpolation method based on rainfall data collected from 1976 to 2010 (Chang et al., 2016). Because these factors are not directly affected by DEM data, they are named non-DEM-based factors. To unify the data format, we resampled all DEM-based and non-DEM-based factors to a 30 m resolution that is consistent with the DEM in ArcGIS Pro and projected them to WGS 1984 UTM Zone 49N (Figure 4 a-f).

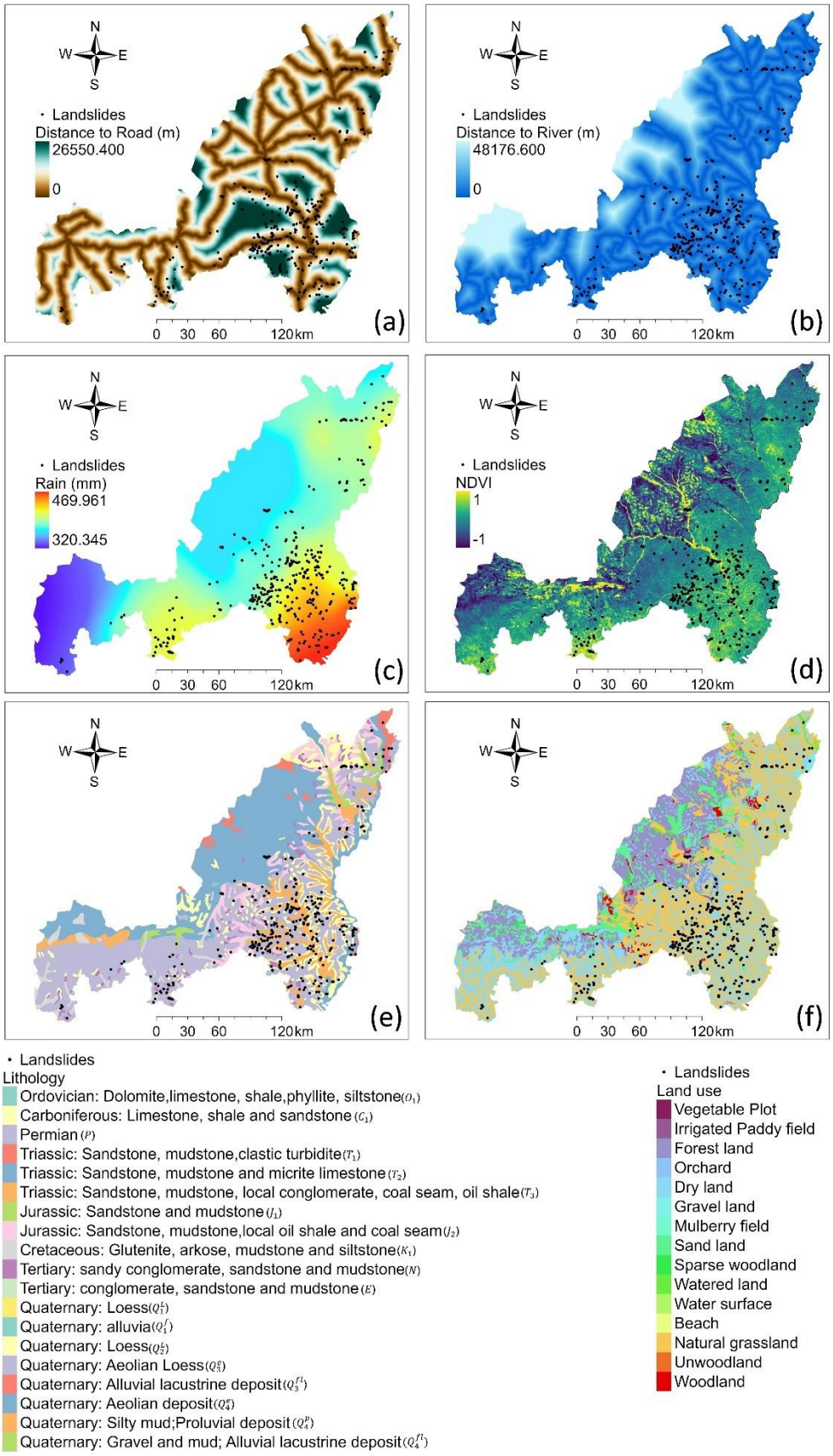

Figure 4. Non-DEM-based factors: (a) distance to road, (b) distance to river, (c) rain, (d) NDVI, (e) lithology, and (f) land use

3. Methodology

3.1 Frequency Ratio for data preparation

Some Non-DEM-based factors are recorded in textual form, such as lithology and land use, which need to be converted to numerical form before being input into ML. FR is chosen to calculate their value, which is represented as "data preparation" in the flowchart (Figure 2), and the calculation method is shown in Equation (1)

$$FR_i = \frac{NoL_i / \sum NoL_i}{NoC_i / \sum NoC_i} \qquad (1)$$

where NoL_i represents the number of landslides in subclass i and $\sum NoL_i$ represents the total number of landslides in all subclasses, i.e., the total number of landslides. $NoL_i / \sum NoL_i$ represents the landslide occupancy rate of subclass i, NoC_i represents the number of pixels contained in subclass i, and $\sum NoC_i$ represents the total number of pixels contained in all subclasses, i.e., the total number of pixels in the entire image. $NoC_i / \sum NoC_i$ represents the area occupancy rate of subclass i. The calculated FR_i represents the frequency ratio, indicating the contribution of subclass i to the occurrence of landslides and whether the landslides are more likely to occur in this subclass. If the calculated FR_i is greater than 1, it indicates that landslides are more closely related to subclass i in this factor and more likely to occur. Conversely, if FR_i is less than 1, it indicates a lower correlation.

3.2 Machine learning

3.2.1 KNN

KNN (K-Nearest Neighbors) is a supervised machine learning algorithm that is relatively simple yet effective. It is particularly suitable for landslide susceptibility mapping (LSM) due to its adaptability to unevenly distributed al., 2020). The underlying principle of the KNN algorithm is that objects with similar features are closer in the feature space. It relies on the majority voting of an object's neighbors to assign it to the class that is most common among its k nearest neighbors.
There are several important steps involved in the KNN algorithm. The first step is to calculate the distance between points using various methods, such as Euclidean, Manhattan, or Minkowski distance. In this particular article, the Euclidean distance measure is chosen, as shown in Equation 3:

$$d(x_1, y_1) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (3)$$

The next step is to select the optimal value for k, which is determined through cross-validation. In this particular article, after performing cross-validation, the value of k equal to 3 is selected as the best choice for the KNN algorithm.

3.2.2 RF

Random Forest (RF) is another supervised classification algorithm proposed by Breiman (2001). It consists of an ensemble of decision trees, where bagging is applied to the tree learners in order to reduce variance without significantly increasing bias. RF has been successful in avoiding overfitting issues commonly associated with individual decision trees and has demonstrated excellent performance in various classification tasks. Although the results of RF models have low interpretability and are often considered as "black-box" models (Merghadi et al., 2020), they are widely used in Landslide Susceptibility Mapping (LSM) (Sun et al., 2021; Zhou et al., 2021) and are considered to be one of the most robust methods available.

The use of RF in LSM helps in effectively handling the complex relationships between landslide occurrence and the multiple factors that influence it. These factors can include topography, soil properties, land cover, and various other geological and environmental variables. RF's ability to capture these intricate relationships and generate accurate landslide susceptibility maps make it a popular choice in the field of LSM.

3.3 Training and testing

In this study, random sampling is used to divide all landslides into training and testing sets, and cross-validation is performed using two sampling strategies. The first strategy (S1) split all landslide samples into a 70% training set and a 30% testing set, with 284 for training and 123 for testing. The second strategy (S2) used a 67% training set and a 33% testing set, with 271 for training and 136 for testing. In this study, confusion matrix is used to assess and compare the performance of trained results, including various indicators calculated using confusion matrices, such as accuracy, precision, recall, F-measures(F1). The calculation methods for these indicators are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{Precision * Recall}{Presision + Recall}$$

Where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative, which respectively means correctly classified landslides and non-landslides, misclassified landslides and non-landslides.
To evaluate the quality of the generated landslide susceptibility maps, the receiver operating characteristic (ROC) curve is introduced, which is commonly used in LSM evaluations and performs well in assessing landslide susceptibility map performance. To quantify the quality of the mapping performance, the area under the curve (AUC)

is further introduced, which quantifies the accuracy of the mapping result by calculating the area under the ROC curve. When the AUC is closer to 1, it indicates higher mapping performance; the closer it is to 0, the poorer the performance. (Pourghasemi et al., 2012a). This indicator is considered as the most important one in the research. In this research, AUC is called success rate for training, while it is used for testing AUC is called prediction rate.

4. Results

4.1 Area comparison of DEM sources

The differences between DEM data sources are propagated to their derived DEM-based factors and amplified to some extent. The differences are significant when comparing these factors between DEMs (Figure 5).

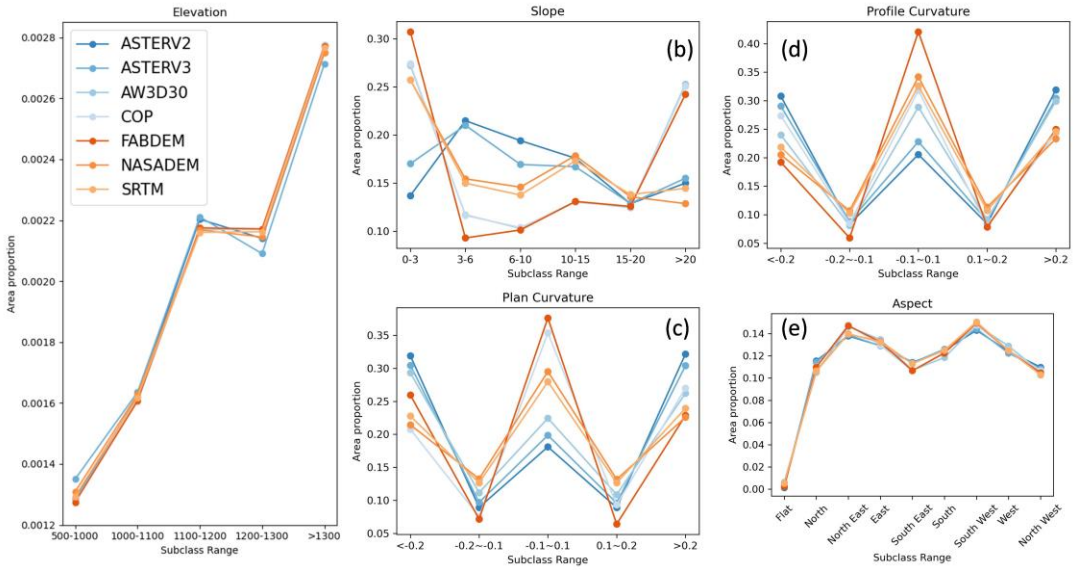

Figure 5 Area proportion of subclasses for different DEMs, (a) elevation, (b) slope, (c) plan curvature, (d) profile curvature, and (e) aspect

The most fundamental factor, elevation, exhibits very small differences. ASTERV3 has the highest proportion of the area in the 500~1000 m subclass and the lowest proportion in the 1200~1300 m subclass, as shown in Figure 5(a). ASTERV3 has a visible difference compared with others that are relatively close.
For aspect, Figure 5(b) shows that the differences among all DEMs in each subclass are not significant, with all of them being approximately 1%. The differences mainly focus on the northeast and southeast classes. COP has the largest area percentage in the northeast class, at 14.78%, which is higher than the smallest, ASTERV2, at 13.78%. In the southeast class, ASTERV2 is the largest, while COP is the smallest, with a difference of approximately 1%.

Slope has the largest difference in proportion and trends against subclasses for these

DEMs. According to the similarity of trends between DEMs, these DEMs can be divided into three groups. The first group is NASADEM and SRTM, accounting for a relatively high area of approximately 30% in the subclass of 0~3° and then decreasing and stabilizing at approximately 15%. The second group is ASTERV2 and ASTERV3, which have a smaller proportion in the subclass of 0~3° compared to other groups, only approximately 15%. Then, the proportion increases in the subclass of 3~6°, reaching 20%. After that, as the slope increases, the area proportion decreases and increases from 12% to approximately 15% in subclasses of >20°. The last group is FABDEM, COP and AW3D30, which have a high proportion in the first and last subclass, both exceeding 25%, and a relatively small proportion in the middle subclasses. Among them, AW3D30 and COP are almost identical, making it difficult to distinguish them in the figure.

Similarly, DEMs can be divided into several groups based on the similarity of plan curvature and profile curvature trends, some of which are the same as slope groups (Figure 5(d) and 4(e)). The first group is NASADEM and SRTM, The first group is a and b, which always maintain high similarity. Their area proportion is at a moderate level in the subclass of <-0.2 m^{-1} and >0.2 m^{-1}. For plan curvature, the area proportion of the subclass <-0.2 m^{-1} is approximately 22%, while that of the subclass >0.2 m^{-1} is approximately 24%. For profile curvature, the area proportion of the subclass <-0.2 m^{-1} is approximately 21%, while that of the subclass >0.2 m^{-1} is approximately 25%. Their area proportion in the subclass of -0.1~0.1 m^{-1} is 27% for plan curvature and 33% for profile curvature. The second group is ASTERV2 and ASTERV3, which have the highest area proportion in both the first and last subclasses, i.e., approximately 32%. The area proportion in the middle subclass is the lowest, with an area proportion of approximately 18% for plan curvature and approximately 21% for profile curvature. AW3D30 is not significantly different from them in terms of plan curvature and can be classified as the same class, but the difference in profile curvature increases, especially in the class of -0.1 m^{-1}~0.1 m^{-1} Similar to the slope, COP and FABDEM maintain very good similarity in the Plan curvature, and have a very high proportion in the subclass of -0.1~0.1 m^{-1} compared to other DEMs, exceeding 35%. However, in the profile curvature, their differences become larger, with area proportion of subclasses -0.1~0.1 m^{-1} over 40% for FABDEM, while only 31.79% for COP.

4.2 KNN

Table 2 shows the results obtained from different DEMs considering 70% of the data for training. The largest precision difference between the DEMs is observed between COP and AW3D30. COP has a precision of 0.9039, while AW3D30 has a precision of 0.7402, resulting in a significant difference of 0.1637. On the other hand, the smallest difference is observed in the recall values. AW3D30 has a recall of 0.8768, while ASTERV2 and NASADEM both have a recall of 0.9049. The difference between them is only 0.0281. When considering the success rate, COP has the largest value of 0.9658.

Conversely, AW3D30 has the smallest success rate of 0.8603, resulting in a significant difference of 0.1055.

Table 2 KNN training with 70% data (S1)

	ASTERV2	*ASTERV3*	*AW3D30*	*COP*	*FABDEM*	*NASADEM*	*SRTM*	*Difference*	*Rank of Difference*
Accuracy	*0.7958*	*0.7817*	*0.7799*	*0.8996*	*0.7870*	*0.8116*	*0.7905*	*0.1197*	*2*
Precision	*0.7642*	*0.7439*	*0.7402*	*0.9039*	*0.7523*	*0.7706*	*0.7508*	*0.1637*	*1*
Recall	*0.9049*	*0.8944*	*0.8768*	*0.8944*	*0.8873*	*0.9049*	*0.9049*	*0.0281*	*5*
F1	*0.8073*	*0.7974*	*0.7967*	*0.8991*	*0.8007*	*0.8249*	*0.8059*	*0.1024*	*4*
Success rate	*0.8646*	*0.8671*	*0.8603*	*0.9658*	*0.8669*	*0.8782*	*0.8742*	*0.1055*	*3*
RANK (success rate)	*6*	*4*	*7*	*1*	*5*	*2*	*3*	*-*	*-*

Table 3 presents the results obtained when the remaining 30% of the data is used for testing. The largest difference is observed in the recall values. COP achieves the highest recall value of 0.8943, while ASTERV2 has the lowest recall value of 0.7154, which has a significant difference of 0.1789. In contrast, the smallest difference is found in the prediction rate. COP exhibits an prediction rate of 0.8701, whereas ASTERV3 has a prediction rate of 0.8299, with a smaller difference of 0.0402.

Table 3 KNN testing with 30% data (S1)

	ASTERV2	*ASTERV3*	*AW3D30*	*COP*	*FABDEM*	*NASADEM*	*SRTM*	*Difference*	*Rank of Difference*
Accuracy	*0.7480*	*0.7561*	*0.7846*	*0.8415*	*0.7398*	*0.7846*	*0.7724*	*0.1017*	*2*
Precision	*0.7226*	*0.7405*	*0.7652*	*0.8088*	*0.7252*	*0.7500*	*0.7445*	*0.0862*	*4*
Recall	*0.7154*	*0.7724*	*0.8293*	*0.8943*	*0.8130*	*0.8211*	*0.7967*	*0.1789*	*1*
F1	*0.7615*	*0.7638*	*0.7922*	*0.8494*	*0.7480*	*0.7985*	*0.7846*	*0.1014*	*3*
Prediction rate	*0.8300*	*0.8299*	*0.8528*	*0.8701*	*0.8433*	*0.8515*	*0.8457*	*0.0402*	*5*
RANK (prediction rate)	*6*	*7*	*2*	*1*	*5*	*3*	*4*	*-*	*-*

According to Table 4, when 67% of the data is used for training, the largest difference in precision among the DEMs is observed. COP achieves a precision of 0.8768, while ASTERV2 has a precision of 0.7908, resulting in a difference of 0.0860. The indicator with the smallest difference is recall. NASADEM exhibits the highest recall value of 0.9262, while FABDEM has the lowest recall value of 0.8893, resulting in a difference of 0.0369. In terms of the success rate, COP achieves the highest value of 0.9644, while both ASTERV2 and FABDEM have the lowest success rate values of 0.9071, resulting in a difference of 0.0573.

Table 4 KNN training with 67% data (S2)

	ASTERV2	*ASTERV3*	*AW3D30*	*COP*	*FABDEM*	*NASADEM*	*SRTM*	*Difference*	*Rank of Difference*
Accuracy	*0.8284*	*0.8321*	*0.8358*	*0.8948*	*0.8284*	*0.8413*	*0.8376*	*0.0664*	*2*
Precision	*0.7908*	*0.8041*	*0.8116*	*0.8768*	*0.8090*	*0.7994*	*0.7980*	*0.086*	*1*
Recall	*0.9225*	*0.9077*	*0.8967*	*0.9188*	*0.8893*	*0.9262*	*0.9188*	*0.0369*	*5*
F1	*0.8388*	*0.8395*	*0.8419*	*0.8973*	*0.8336*	*0.8517*	*0.8478*	*0.0637*	*3*

Success rate	*0.9071*	*0.9104*	*0.9073*	*0.9644*	*0.9071*	*0.9141*	*0.9133*	*0.0573*	*4*
RANK (success rate)	*7*	*4*	*5*	*1*	*6*	*2*	*3*	*-*	*-*

In Table 5, the largest difference is observed in precision, with COP having the highest precision value of 0.8261 and ASTERV2 having the lowest precision value of 0.7857. The difference between COP and ASTERV2 is 0.0404, indicating a noticeable difference in precision. On the other hand, the smallest difference of 0.0207 is found in prediction rate. NASADEM achieves the highest prediction rate of 0.9104, while ASTERV2 has the lowest prediction rate of 0.8897.

Table 5 KNN testing with 33% data (S2)

	ASTERV2	*ASTERV3*	*AW3D30*	*COP*	*FABDEM*	*NASADEM*	*SRTM*	*Difference*	*Rank of Difference*
Accuracy	*0.7941*	*0.7978*	*0.7941*	*0.8309*	*0.8088*	*0.8272*	*0.8051*	*0.0368*	*4*
Precision	*0.7857*	*0.7956*	*0.7985*	*0.8261*	*0.8134*	*0.8112*	*0.8029*	*0.0404*	*1*
Recall	*0.8088*	*0.8162*	*0.8309*	*0.8382*	*0.8309*	*0.8456*	*0.8309*	*0.0368*	*3*
F1	*0.7971*	*0.7985*	*0.7926*	*0.8321*	*0.8074*	*0.8315*	*0.8059*	*0.0395*	*2*
Prediction rate	*0.8897*	*0.8917*	*0.8973*	*0.8920*	*0.8987*	*0.9104*	*0.9075*	*0.0207*	*5*
RANK (prediction rate)	*7*	*6*	*4*	*5*	*3*	*1*	*2*	*-*	*-*

Precision is the indicator that best reflects the differences between the DEMs in most cases, as it consistently has the highest difference rankings. This suggests that precision is a reliable indicator. Precision consistently leads in performance, followed by accuracy. Recall and F1 exhibit moderate and fluctuating rankings, indicating their varying performance across scenarios. Success rate and prediction rate is considered as the most important indicators for evaluation. In most cases, COP has a very prominent advantage compared to other DEMs. Additionally, ASTERV2 and ASTERV3 have poor performance in most cases.

When comparing the results of two sampling methods, it is observed that most indicators of sampling method S2 are greater than those of S1. This consistent pattern indicates that the choice of sampling method has a stable impact on the results. Generally, S2 yields better results compared to S1. However, for COP, the situation is different. Most indicators of S1 are greater than those of S2, indicating that for COP, a larger proportion of the training dataset is required to achieve better results. This suggests that the performance of COP is more sensitive to the size of the input training dataset, and a larger dataset produces more favorable outcomes for this model.

When examining the outcomes of various sampling methods on both the training and testing datasets, a consistent trend emerges: the performance on the training set

consistently surpasses that on the testing set. This outcome is reasonable and expected in machine learning scenarios. The superior performance of the training dataset indicates that the model has successfully learned the desired features and patterns from this dataset. Through the training process, the model adjusted its parameters and optimized its performance to fit the training data. Figure 6 shows the landslide susceptibility map of KNN, and we can see that most landslide points are distributed in areas with higher index.

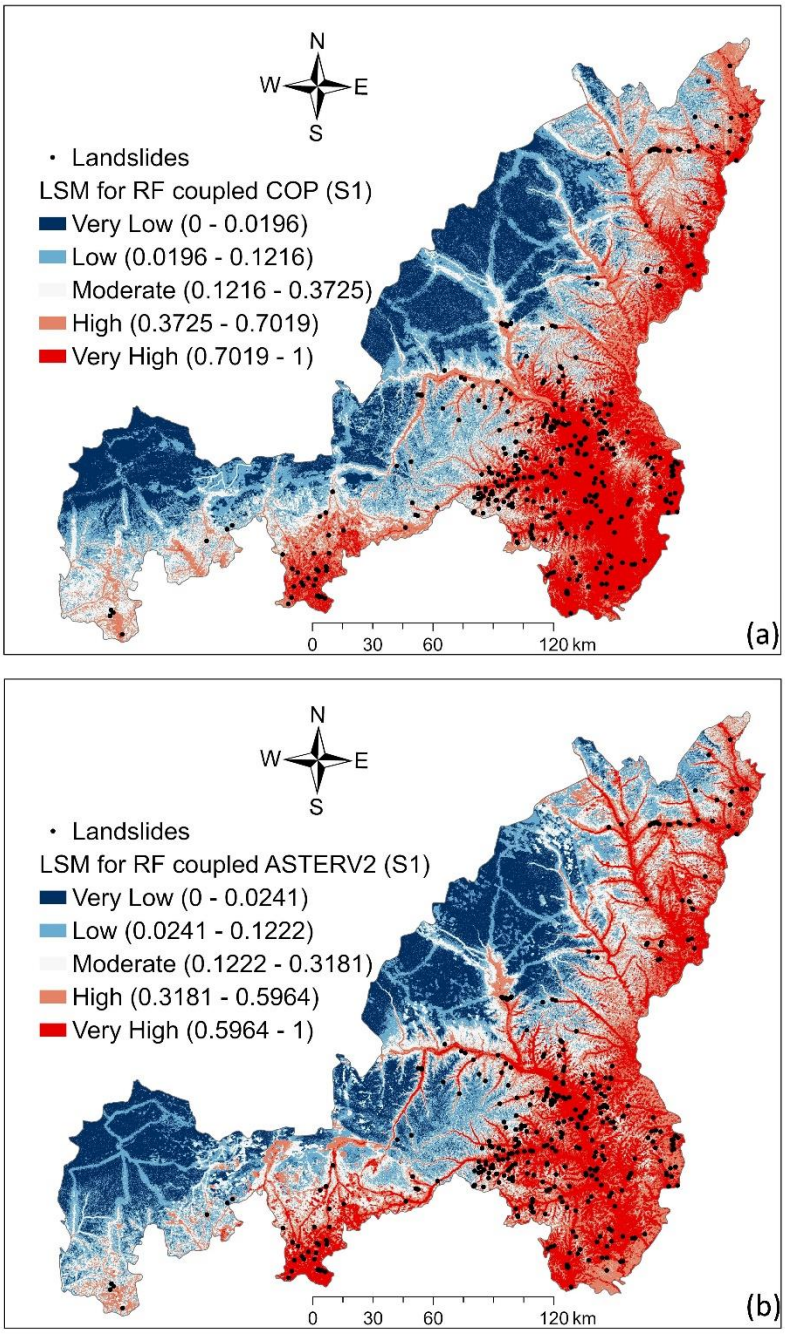

Figure 6 Landslide susceptibility map of KNN (a: COP with S2; b: ASTERV2 with S2)

4.3 RF

For RF, regardless of the sampling method, the success rate of the training set is 1, which means all samples can be correctly classified into their respective categories. So, we only show the prediction rate of the testing dataset in table 6 and 7.

Table 6 RF testing with 30% data (S1)

	ASTERV2	ASTERV3	AW3D30	COP	FABDEM	NASADEM	SRTM	Difference	Rank of Difference
Accuracy	0.8293	0.8333	0.8333	0.8211	0.8415	0.8293	0.8211	0.0204	3
Precision	0.8092	0.8203	0.8154	0.7926	0.8231	0.8045	0.7970	0.0305	1
Recall	0.8618	0.8537	0.8618	0.8699	0.8699	0.8699	0.8618	0.0162	5
F1	0.8346	0.8367	0.8379	0.8295	0.8458	0.8359	0.8281	0.0177	4
Prediction rate	0.8996	0.9047	0.9149	0.9254	0.9125	0.9145	0.9123	0.0258	2
RANK (Prediction rate)	7	6	2	1	4	3	5	-	-

In the scenario where 30% of the data is used for testing, the differences in performance among the digital elevation models (DEMs) can be observed in terms of precision, recall, and prediction rateprediction rate. FABDEM achieves the highest precision value of 0.8231, while COP has a slightly lower precision value of 0.7926. The difference between FABDEM and COP in terms of precision is 0.0305. COP, FABDEM, and NASADEM all exhibit the largest recall value of 0.8699. On the other hand, ASTERV3 has the smallest recall value of 0.8537. The difference between ASTERV3 and the other DEMs in terms of recall is 0.0162. COP achieves the largest prediction rateprediction rate of 0.9254, suggesting its stronger ability to distinguish between positive and negative instances. On the other hand, ASTERV2 has the smallest prediction rateprediction rate of 0.8996. The difference between COP and ASTERV2 in terms of prediction rateprediction rate is 0.0258.

Table 7 RF testing with 33% data (S2)

	ASTERV2	ASTERV3	AW3D30	COP	FABDEM	NASADEM	SRTM	Difference	Rank of Difference
Accuracy	0.8603	0.8713	0.8603	0.8860	0.8493	0.8529	0.8529	0.0367	3
Precision	0.8657	0.8686	0.8551	0.8671	0.8467	0.8582	0.8529	0.0219	4
Recall	0.8529	0.8750	0.8676	0.9118	0.8529	0.8456	0.8529	0.0662	1
F1	0.8593	0.8718	0.8613	0.8889	0.8498	0.8519	0.8529	0.0391	2
Prediction rate	0.9361	0.9411	0.9356	0.9461	0.9371	0.9338	0.9347	0.0123	5
RANK (Prediction rate)	4	2	5	1	3	7	6	-	-

When using 33% of the data for testing, the differences in performance among the DEMs can be observed in terms of recall and prediction rateprediction rate. COP achieves the largest recall value of 0.9118, indicating its ability to correctly identify

positive instances. On the other hand, ASTERV2, AW3D30, and SRTM perform relatively lower, with a recall value of 0.8529. The difference between these DEMs in terms of recall is 0.0662. The differences in prediction rateprediction rate among the DEMs are relatively smaller compared to other indicators. COP has the largest prediction rateprediction rate of 0.9461. NASADEM, on the other hand, has the smallest prediction rateprediction rate of 0.9338. The difference between COP and NASADEM in terms of prediction rateprediction rate is 0.0123.
Overall, comparing the two sampling methods, a similar conclusion can be drawn as before - the performance of sampling method S2 is better than that of S1. However, it is noteworthy that for the RF model, the recall of ASTERV2, FABDEM, NASADEM, and SRTM is greater with S1 compared to S2.

Based on the findings, it is observed that the performance of RF is generally superior to the KNN model for most DEMs. This indicates that the RF has better predictive capabilities and achieves higher accuracy compared to KNN, except for the case of COP. The reason behind the exception of COP could be attributed to the quality and characteristics of this particular DEM. COP might already possess high-quality data and provide sufficient information to accurately predict landslide susceptibility. As a result, the improvement gained by using RF instead of KNN may not be significant for COP. In other words, COP itself might be performing at a high level, minimizing the added value of RF for this specific DEM.

Figure 7 shows the landslide susceptibility map of RF. It is shown that most landslides are located in highly susceptibility areas.

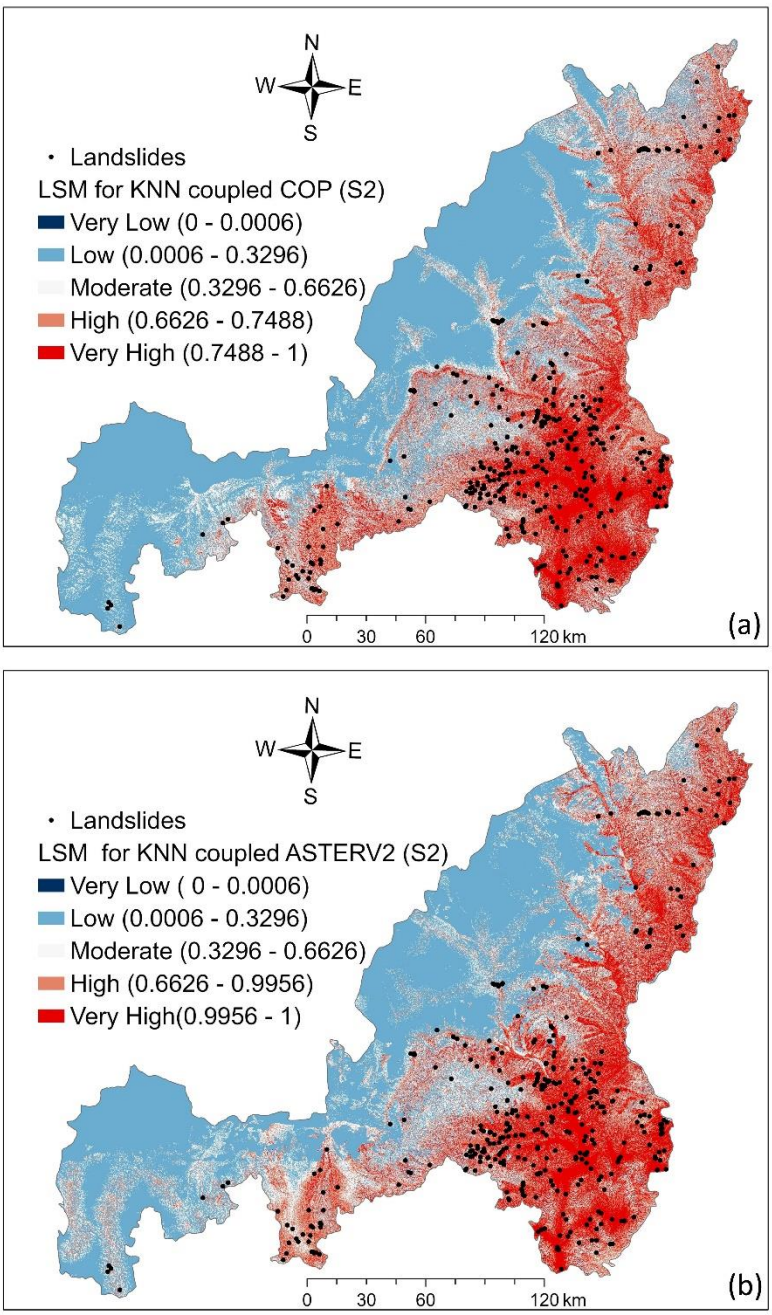

Figure 7 Landslide susceptibility of RF (a: COP with S1; b: ASTERV2 with S1)

5. Discussion

5.1 Uncertainty analysis of sampling

The paper uses multiple data sampling methods to verify the stability of the result. The main experiment uses the method of using all landslides for training (100%). Subsequently, two groups of experiments were conducted using the 70% training and 30% testing (S1) and the 67% training and 33% testing (S2) methods.

As is mentioned early, two sampling methods have been compared, and whether it is

from the perspective of evaluation indicators or numerical ranges, the sampling methods of S2 are more advantageous in most cases. However, as the sampling method changes, the rank of DEMs also changes. The ranking problem and the stability of DEMs need to be further elaborated.

To further quantify the stability, i.e., its similarity to the main experiment results, a result matrix is constructed for each group of experiments, with the number of rows and columns equal to the number of DEMs, i.e., a 7x7 matrix R. Each pair of DEMs is compared based on their success rate or prediction rateprediction rate rankings, and a 1 is filled in at position (A, B) in matrix R if Rank(A) > Rank(B), and a is -1 otherwise. After obtaining the result matrix R for each group of experiments, they are added up with the result matrix for the main experiment to obtain a comparison matrix C. The comparison matrix may have values of 2, -2, or 0 at each position. The stability was calculated using Eq. (3).

$$Stability = \frac{\sum_{A,B}[C(A,B)=2]+\sum_{A,B}[C(A,B)=-2]}{\sum_{A,B}[C(A,B)=2]+\sum_{A,B}[C(A,B)=-2]+\sum_{A,B}[C(A,B)=0]} \qquad (3)$$

In this equation, $\sum_{A,B}[C(A,B)=2]$ represents the number of positions in comparison matrix C, where the value is 2, while $\sum_{A,B}[C(A,B)=-2]$ represents the number of positions in comparison matrix C, where the value is -2. The sum of these two values indicates the number of pairs for which the order is the same in both experiments, A and B. The denominator represents the total number of compared DEM pairs. The ratio of these two values represents the proportion of pairs whose order is stable among all pairs.

There are a total of six possible combinations of the model (KNN, RF) and sampling method (S1, S2, training, testing), which are cross compared and the stability results obtained are recorded in Table 8.

Table 8 Stability

	KNN S1(Training)	*KNN S1(Testing)*	*KNN S2(Training)*	*KNN S2(Testing)*	*RF S1(Testing)*	*RF S2(Testing)*
KNN S1(Training)	-	*0.7143*	*0.8776*	*0.7143*	*0.7143*	*0.6327*
KNN S1(Testing)	-	-	*0.7551*	*0.6735*	*0.9184*	*0.5102*
KNN S2(Training)	-	-	-	*0.6735*	*0.7551*	*0.5102*
KNN S2(Testing)	-	-	-	-	*0.6735*	*0.3469*
RF S1(Testing)	-	-	-	-	-	*0.5918*
RF S2(Testing)	-	-	-	-	-	-

The stability range of sorting is 0.3469-0.9184, with the largest results being KNN 30% data for testing combining with RF 30% data for testing. This indicates that on a 30% testing dataset, the model has little impact and the results obtained are very close. The biggest difference is that KNN 33% is used for testing combining with RF with 33% data for testing. These two used the same testing dataset, but the similarity of the results was very small and the stability of the results was poor, indicating that model had a significant impact on this dataset.

The results of KNN are compared internally, with a ranking stability range of 0.6735-0.8776, indicating overall stability. However, comparing internally between RF, the stability is only 0.5918, indicating that the RF model has relatively low stability and is sensitive to samples and sampling methods.

5.2 Assessment of DEM sources for LSM

Previous research has extensively studied the absolute accuracy of open-source DEMs by comparing the elevation values with ground control points at the national scale. The resulting errors are recorded in terms of root mean square error (RMSE) values (Buckley et al., 2020; Gesch et al., 2016; Gonzalez-Ollauri and Mickovski, 2017; Hirt et al., 2010; Mouratidis et al., 2010; Tachikawa et al., 2011; Uuemaa et al., 2020), which are compiled in Table 1 as "Precision" values.

The accuracy of the DEM itself may not necessarily be completely consistent with the mapping performance, but the similarity between DEMs is often reflected in the mapping process and results, such as SRTM and NASADEM, ASTERV2 and ASTERV3. Wang et al. (2021) found that NASADEM had a much lower error than ASTERV3. This study found that NASADEM has the best mapping performance among all DEMs. Uuemaa et al. (2020) found that SRTM and NASADEM both performed well, with only a slight improvement in NASADEM over SRTM. This study found that NASADEM and SRTM are very similar for the area proportion and FR range, and the mapping results are also very close. Elkhrachy (2018) found that in Saudi Arabia, ASTERV2 is slightly more accurate than SRTM, but both had errors far below the official threshold. However, the mapping performance of ASTERV2 is not as high as that of SRTM in this study, and its results are very similar to ASTERV3, both of which are poorer compared to other DEMs. Uuemaa et al. (2020) also found that AW3D30 has the highest accuracy compared to ASTERV2, MERIT, TanDEM-X, SRTM, and NASADEM, but it ranks only fourth in the mapping performance. Guth and Geoffroy (2021) found that COP has the highest accuracy compared to ALOS, ASTER, NASA, and SRTM. Li et al. (2022) found that COP has similar vertical accuracy to NASADEM and AW3D30, and COP displays the most detailed terrain information. Hawker et al. (2022) found that FABDEM is able to represent smaller hydrological features than other global DEMs.

The research findings indicate that there is a significant correlation between the results of landslide susceptibility mapping (LSM) and the digital elevation models (DEMs) used. Among the DEMs evaluated, COP consistently demonstrated excellent performance and vertical accuracy in many cases. It has proven to be a reliable choice for LSM. On the other hand, ASTERV2 and ASTERV3, despite being commonly used DEMs, did not perform well in LSM. These DEMs share the same source and exhibit high similarity, but their performance is not satisfactory, suggesting that they should be avoided for landslide susceptibility mapping purposes. In contrast, NASADEM and SRTM, also sourced differently, have shown high similarity and performed well in LSM. Despite their similar origin, these DEMs maintain their individual strengths. It is worth noting that FABDEM, which is the result of optimizing hydrological relationships through machine learning applied to COP, did not perform as well as COP in LSM. This suggests that the original COP DEM is more reliable and preferable for LSM applications.

5.3 Comparison of DEM selection among model, factors and subclass division

In LSM, the choice of model, factors, subclass division, and DEM selection all have some influence on the mapping results.

Model selection has a significant impact on prediction rateprediction rate. Fang et al. (2021) compared four models and found a difference in prediction rateprediction rate of approximately 7% between the highest and lowest accuracies. Kavzoglu et al. (2015) compared three multivariate methods and three bivariate methods and found that the multivariate methods were better than the bivariate methods, with support vector regression performing the best and weight of evidence the weakest, with a difference in accuracy of approximately 13%. Other studies (Choi et al., 2012; Nohani et al., 2019; Park et al., 2013) have also shown that each model has its advantages and disadvantages, but the choice of model can result in a difference of 5%-20% in accuracy.

Similarly, the combination of selected factors can also affect the result. Kavzoglu et al. (2015) found that the number and combination of factors affect the accuracy, with a plateau in accuracy observed after eight factors. Slope, lithology, and distance to the water system are relatively effective factors, and the difference in the accuracy between the best and worst factors is approximately 11%. Micheletti et al. (2014) found that the choice of factors could result in a difference of up to 20% in results. Wang et al. (2015) found that the combination of selected factors can produce a difference of up to 7%.

Subclass division is the most subjective part of LSM, and different subclass division criteria can also affect the mapping results. Yan et al. (2019) found that different subclass division criteria can result in a difference in the accuracy of 2-3%.

In this research, it was found that different models and sampling methods have different effects on the results. For KNN, this difference has reached 10.55%, with a minimum of 2.07%. For RF, the maximum difference is 2.58% and the minimum is 1.23%.

6. Conclusion

In this research, when comparing the area proportions of different DEMs, significant differences were observed. The study noted that the differences in elevation and aspect were not significant, with only around a 1% variation. However, when considering slope, plan curvature, and profile curvature, the differences were much more pronounced, ranging from 15% to 20%. This suggests that these terrain attributes are more sensitive to variations in the DEMs. Furthermore, the study identified three groups of DEMs based on their similarity. The first group consists of ASTERV2 and ASTERV3, which exhibit high similarity to each other. The second group includes NASADEM and SRTM, which also share a significant similarity. The last group comprises FABDEM, COP, and AW3D30, displaying similarity among themselves.

Two models, KNN and RF, were utilized for landslide susceptibility prediction and mapping. The performance of these models was evaluated using various indicators, including precision, accuracy, recall, F1 score, and prediction rateprediction rate. For KNN, it was observed that recall exhibited the greatest difference among all indicators. The recall values ranged from 0.7154 to 0.8943, showing a difference of 0.1789. When compared to sampling method S2, the recall difference was 0.0386. Additionally, accuracy exhibited the second-largest difference, ranging from 0.7398 to 0.8415 with a difference of 0.1017. When use the sampling method S2, the difference decreases to 0.0368. Regarding precision, there were differences of 0.0862 and 0.0404 for S1 and S2, respectively. Prediction ratePrediction rate, being the indicator that best represents mapping performance, showed differences of 0.0402 and 0.0207, respectively. For RF, the differences in all indicators decreased to a certain extent. With sampling method S1, the difference in accuracy, precision, recall, and F1 decreased to 0.0204, 0.0305, 0.0162, and 0.0177, respectively. When using sampling method S2, the differences in these indicators were 0.0367, 0.0219, 0.0662, and 0.0391, respectively.

In general, COP has the best mapping performance, due to the highest prediction rateprediction rate of 0.8701, 0.9254 and 0.9461, for KNN with S1 and for RF with S1 and S2, respectively. While ASTERV2 is the worst with lowest prediction rateprediction rate of 0.8897 and 0.8996, for KNN with S2 and for RF with S1.

In this research, it is also found that the ranking results between DEMs have strong randomness. In the process of comparing the ranking results with each other, the

highest stability reached 0.9184, but the lowest was only 0.3469.

References

Arumugam, T., Kinattinkara, S., Velusamy, S., Shanmugamoorthy, M. and Murugan, S. (2023) GIS based landslide susceptibility mapping and assessment using weighted overlay method in Wayanad: A part of Western Ghats, Kerala. Urban Climate, 49, 101508.

Conforti, M., Pascale, S., Robustelli, G. and Sdao, F. (2014) Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). CATENA, 113, 236–250.

Gonzalez-Ollauri, A. and Mickovski, S.B. (2017) Hydrological effect of vegetation against rainfall-induced landslides. Journal of Hydrology, 549, 374–387.

Guo, W., Chen, Z., Wang, W., Gao, W., Guo, M., Kang, H., Li, P., Wang, W. and Zhao, M. (2020) Telling a different story: The promote role of vegetation in the initiation of shallow landslides during rainfall on the Chinese Loess Plateau. Geomorphology, 350, 106879.

Huang, G., Zheng, M. and Peng, J. (2021) Effect of Vegetation Roots on the Threshold of Slope Instability Induced by Rainfall and Runoff. Geofluids, e6682113.

Mohammady, M., Pourghasemi, H.R. and Pradhan, B. (2012) Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. Journal of Asian Earth Sciences, Geological Anatomy of East and South Asia, 61, 221-236.

Poudyal, C, Chang, C., Oh, H. and Lee, S. (2010) Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. Environ Earth Sci, 61, 1049–1064.

Sun, D., Wen, H., Wang, D. and Xu, J. (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology, 362, 107201.

Vakhshoori, V. and Zare, M. (2016) Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. Geomatics, Natural Hazards and Risk, 7, 1731–1752.

Florinsky, I.V., Skrypitsyna, T.N. and Luschikova, O.S. (2018) Comparative accuracy of the AW3D30 DSM, ASTER GDEM, and SRTM1 DEM: A case study on the Zaoksky testing ground, Central European Russia. Remote Sensing Letters, 9, 706–714.

Jacobs, L., Dewitte, O., Poesen, J., Sekajugo, J., Nobile, A., Rossi, M., Thiery, W. and Kervyn, M. (2018) Field-based landslide susceptibility assessment in a data-scarce environment: the populated areas of the Rwenzori Mountains. Natural Hazards and Earth System Sciences, 18, 105-124.

Meena, S.R. and Gudiyangada Nachappa, T. (2019) Impact of Spatial Resolution of Digital Elevation Model on Landslide Susceptibility Mapping: A Case Study in Kullu Valley, Himalayas. Geosciences, 9, 360.

Tian, Y., XiaO, C., Liu, Y. and Wu, L. (2008) Effects of raster resolution on landslide susceptibility mapping: A case study of Shenzhen. Sci. China Ser. E-Technol. Sci., 51,

188-198.
Abu, E.S., Ali, S.A. and Pham, Q.B. (2021) Spatial modeling and susceptibility zonation of landslides using random forest, naïve bayes and K-nearest neighbor in a complicated terrain. Earth Sci Inform, 14:1227–43. https://doi.org/10.1007/s12145-021-00653-y.
Aditian, A., Kubota, T. and Shinohara, Y. (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology, 318, 101-111.
Ado, M., Amitab, K., Maji, A.K., Jasińska, E., Gono, R., Leonowicz, Z. and Jasiński, M. (2022) Landslide Susceptibility Mapping Using Machine Learning: A Literature Survey. Remote Sensing, 14, 3029.
Althuwaynee, O.F., Pradhan, B. and Lee, S. (2012) Application of an evidential belief function model in landslide susceptibility mapping. Computers & Geosciences, 44:120–35. https://doi.org/10.1016/j.cageo.2012.03.003.
Arabameri, A., Pradhan, B., Rezaei, K. and Lee, C.-W. (2019) Assessment of Landslide Susceptibility Using Statistical- and Artificial Intelligence-Based FR–RF Integrated Model and Multiresolution DEMs. Remote Sensing, 11,999.
Ayalew, L. and Yamagishi, H. (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology, 65, 15–31.
Batar, A.K. and Watanabe, T. (2021) Landslide Susceptibility Mapping and Assessment Using Geospatial Platforms and Weights of Evidence (WoE) Method in the Indian Himalayan Region: Recent Developments, Gaps, and Future Directions. ISPRS International Journal of Geo-Information, 10, 114.
Breiman, L. (2001) Random Forests. Machine Learning, 45:5–32. https://doi.org/10.1023/A:1010933404324.
Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M. and Brenning, A. (2020) The performance of landslide susceptibility models critically depends on the quality of digital elevation models. Geomatics, Natural Hazards and Risk, 11, 1075–1092.
Broeckx, J., Vanmaercke, M., Duchateau, R. and Poesen, J. (2018) A data-based landslide susceptibility map of Africa. Earth-Science Reviews, 185, 102–121.
Buckley, S. M., Agram, P. S., Belz, J. E., Crippen, R. E., Gurrola, E. M., Hensley, S., ... & Tung, W. W. (2020). NASADEM. National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology: Pasadena, CA, USA.
Chang, X., Shen, B. and Huang, L. (2016) Trend analysis of spatial－temporal change of precipitation in Yulin from 1976 to 2010. Journal of Water Resources and Water, 27, 30–35.
Chen, X. and Chen, W. (2021) GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. CATENA, 196, 104833.
Chen, Z., Ye, F., Fu, W., Ke, Y. and Hong, H. (2020) The influence of DEM spatial resolution on landslide susceptibility mapping in the Baxie River basin, NW China. Nat Hazards, 101, 853–877.
Choi, J., Oh, H., Lee, H., Lee, C. and Lee, S. (2012) Combining landslide susceptibility

maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. Engineering Geology, 124, 12–23.

Dandabathula, G., Hari, R., Ghosh, K., Bera, A.K. and Srivastav, S.K. (2022) Accuracy assessment of digital bare-earth model using ICESat-2 photons: analysis of the FABDEM. Modeling Earth Systems and Environment, 9, 2677–2694.

Dou, J., Yunus, A.P., Tien Bui, D., Sahana, M., Chen, C.-W., Zhu, Z., Wang, W. and Thai Pham, B. (2019) Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. Remote Sensing, 11, 638.

Elkhrachy, I. (2018) Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city, Saudi Arabia. Ain Shams Engineering Journal, 9, 1807–1817.

Fang, Z., Wang, Y., Peng, L. and Hong, H. (2021) A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. International Journal of Geographical Information Science, 35, 321–347.

Feng, L., Zhang, M., Jin, Z., Zhang, S., Sun, P., Gu, T., Liu, X., Lin, H., An, Z., Peng, J. and Guo, L. (2021) The genesis, development, and evolution of original vertical joints in loess. Earth-Science Reviews, 214, 103526.

Gesch, D., Oimoen, M., Danielson, J. and Meyer, D. 2016) Validation of the ASTER Global Digital Elevation Model Version 3 over the conterminous United States. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B4, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic.

Guth, P.L. and Geoffroy, T.M. (2021) LiDAR point cloud and ICESat-2 evaluation of 1 second global digital elevation models: Copernicus wins. Transactions in GIS, 25:2245–61. https://doi.org/10.1111/tgis.12825.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P. (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31, 181-216.

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C. and Neal, J. (2022) A 30 m global map of elevation with forests and buildings removed. Environmental Research Letters, 17, 024016.

Hirt, C., Filmer, M. and Featherstone, W., 2010. Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9S ver3 digital elevation models over Australia. Australian Journal of Earth Sciences, 57, 337-347.

Hong, H., Liu, J., Bui, D., Pradhan, B., Acharya, T., Pham, B., Zhu, A., Chen, W., Ahmad, B. (2018) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). CATENA, 163, 399-413.

Hong, H., Miao, Y., Liu, J. and Zhu, A. (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. CATENA, 176, 45–64.

Hua, Y., Wang, X., Li, Y., Xu, P. and Xia, W. (2021) Dynamic development of landslide

susceptibility based on slope unit and deep neural networks. Landslides, 18, 281-302.
Huang, W., Ding, M., Li, Z., Yu, J., Ge, D., Liu, Q. and Yang, J. (2023) Landslide susceptibility mapping and dynamic response along the Sichuan-Tibet transportation corridor using deep learning algorithms. CATENA, 222, 106866.
Hussain, M.A., Chen, Z., Wang, R. and Shoaib, M. (2021) PS-InSAR-Based Validated Landslide Susceptibility Mapping along Karakorum Highway, Pakistan. Remote Sensing, 13, 4129.
Ioannidis, C., Xinogalas, E. and Soile, S. (2014) Assessment of the global digital elevation models ASTER and SRTM in Greece. Survey Review, 46, 342–354.
Jian, W., Wang, Z. and Yin, K. (2009) Mechanism of the Anlesi landslide in the Three Gorges Reservoir, China. Engineering Geology, 108, 86-95.
Jing, C., Shortridge, A., Lin, S. and Wu, J. (2014) Comparison and validation of SRTM and ASTER GDEM for a subtropical landscape in Southeastern China. International Journal of Digital Earth, 7, 969-992.
Kavzoglu, T., Kutlug Sahin, E. and Colkesen, I. (2015) Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. Engineering Geology, 192, 101-112.
Khusulio, K. and Kumar, R. (2023) Feasibility assessment of multi-criteria decision making and quantitative landslide susceptibility methods: A case study of Mao-Maram Manipur. J Earth Syst Sci, 132, 56.
Lee, S., Ryu, J., Won, J. and Park, H. (2004) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Engineering Geology, 71, 289-302.
Lee, S., Ryu, J.H., Min, K. and Won, J.S. (2003) Landslide susceptibility analysis using GIS and artificial neural network. Earth Surface Processes and Landforms, 28:1361–76. https://doi.org/10.1002/esp.593.
Li, H., Zhao, J., Yan, B., Yue, L. and Wang, L. (2022) Global DEMs vary from one to another: an evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data. International Journal of Digital Earth, 15:1149–68. https://doi.org/10.1080/17538947.2022.2094002.
Li, L., Lan, H., Guo, C., Zhang, Y., Li, Q. and Wu, Y. (2017) A modified frequency ratio method for landslide susceptibility assessment. Landslides, 14, 727–741.
Lin, Q. and Wang, Y. (2018) Spatial and temporal analysis of a fatal landslide inventory in China from 1950 to 2016. Landslides, 15, 2357–2372.
Mahalingam, R. and Olsen, M. (2016) Evaluation of the influence of source and spatial resolution of DEMs on derivative products used in landslide mapping. Geomatics, Natural Hazards and Risk, 7, 1835–1855.
Merghadi, A., Yunus, A., Dou, J., Whiteley, J., ThaiPham, B., Bui, D., Avtar, R. and Abderrahmane, B. (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225.
Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B. and Bui, D.T. (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207:103225.

https://doi.org/10.1016/j.earscirev.2020.103225.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M. and Kanevski, M.(2014) Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping. Math Geosci, 46, 33-57.

Mouratidis, A., Briole, P. and Katsambalos, K. (2010) SRTM 3" DEM (versions 1, 2, 3, 4) validation by means of extensive kinematic GPS measurements: a case study from North Greece. International Journal of Remote Sensing, 31, 6205–6222.

Neupane, A., Paudyal, K.R., Devkota, K.C. and Dhungana, P. (2023) Landslide susceptibility analysis using frequency ratio and weight of evidence approaches along the Lakhandehi Khola watershed in the Sarlahi District, southern Nepal. Geographical Journal of Nepal, 73–96.

Nohani, E., Moharrami, M., Sharafi, S., Khosravi, K., Pradhan, B., Pham, B.T., Lee, S. and M. Melesse, A. (2019) Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. Water, 11, 1402.

Okolie, C.J. and Smit, J.L. (2022) A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 188:1–29. https://doi.org/10.1016/j.isprsjprs.2022.03.016.

Panchal, S. and Shrivastava, A.Kr. (2022) Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India. Ain Shams Engineering Journal, 13, 101626.

Pardeshi, S.D., Autade, S.E. and Pardeshi, S.S. (2013) Landslide hazard assessment: recent trends and techniques. SpringerPlus, 2, 523.

Park, S., Choi, C., Kim, B. and Kim, J. (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci, 68, 1443–1464.

Peng, J., Hungchou, L., Qiyao, W., Jianqi, Z., Yuxiang, C. and Xinghua, Z. (2014) The critical issues and creative concepts in mitigation research of loess geological hazards. Journal of engineering Geology, 22, 684–691.

Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y. and Ye, R. (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. Geomorphology, 204, 287-301.

Pourghasemi, H., Mohammady, M. and Pradhan, B. (2012a) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. CATENA, 97, 71–84.

Pourghasemi, H.R., Pradhan, B. and Gokceoglu, C. (2012b) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat Hazards 63, 965–996.

Qin, C., Bao, L., Zhu, A., Wang, R. and Hu, X. (2013) Uncertainty due to DEM error in landslide susceptibility mapping. International Journal of Geographical Information Science, 27, 1364–1380.

Rabby, Y., Ishtiaque, A. and Rahman, Md. (2020) Evaluating the Effects of Digital Elevation Models in Landslide Susceptibility Mapping in Rangamati District, Bangladesh. Remote Sensing, 12, 2718.

Rohan, T.J., Wondolowski, N. and Shelef, E. (2021) Landslide susceptibility analysis based on citizen reports. Earth Surface Processes and Landforms, 46:791–803. https://doi.org/10.1002/esp.5064.

Saleem, N., Huq, M., Twumasi, N., Javed, A. and Sajjad, A. (2019) Parameters Derived from and/or Used with Digital Elevation Models (DEMs) for Landslide Susceptibility Mapping and Landslide Risk Assessment: A Review. ISPRS International Journal of Geo-Information, 8, 545.

Sameen, M.I., Pradhan, B., Bui, D.T. and Alamri, A.M. (2020) Systematic sample subdividing strategy for training landslide susceptibility models. CATENA, 187:104358. https://doi.org/10.1016/j.catena.2019.104358.

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M. and Malet, J. (2018) Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. Geomorphology, 301, 10-20.

Shi, J., Chen, L., Shi, J., Fu, B. and Zhang, S. (2003) Regional Characteristics and Driving Mechanism of Land use/Cover Change in Yulin Prefecture, Shaanxi. Scientia Geographica Sinica, 493-498.

Sun, D., Ding, Y., Wen, H. and Zhang, F. (2023) A novel QLattice-based whitening machine learning model of landslide susceptibility mapping. Earth Surface Processes and Landforms. https://doi.org/10.1002/esp.5675.

Sun, D., Xu, J., Wen, H. and Wang, D. (2021) Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. Engineering Geology, 281:105972. https://doi.org/10.1016/j.enggeo.2020.105972.

Sun, J., Wang, T. and Yan, C. (2012) The Relative Roles of Climate Change and Human Activities in Desertification Process:A case study in Yulin,Shaanxi Province,China. Journal of Desert Research, 32, 625-630.

Tachikawa, T., Hato, M., Kaku, M. and Iwasaki, A. (2011) Characteristics of ASTER GDEM version 2. In: 2011 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2011 - 2011 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Vancouver, BC, Canada, pp. 3657–3660.

Uchida, T., Kosugi, K. and Mizuyama, T. (2001) Effects of pipeflow on hydrological process and its relation to landslide: a review of pipeflow studies in forested headwater catchments. Hydrological Processes, 15, 2151–2174.

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M. and Kmoch, A. (2020) Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). Remote Sensing, 12, 3482.

van Westen, C.J., Castellanos, E. and Kuriakose, S.L. (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Engineering Geology, Landslide Susceptibility, Hazard and Risk Zoning for Land Use Planning, 102, 112–131.

Wang, Q., Wang, D., Huang, Y., Wang, Z., Zhang, L., Guo, Q., Chen, W., Chen, W. and Sang, M. (2015) Landslide Susceptibility Mapping Based on Selected Optimal Combination of Landslide Predisposing Factors in a Large Catchment. Sustainability, 7, 16653–16669.

Wang, Y. and Chen, S. (2011) A Survey of Evaluation and Design for AUC Based Classifier. Pattern Recognition and Artificial Intelligence, 24, 64 -71.

Wang, Y., Wang, L. and He, F. (2021) Comparative Study of Data Accuracy from NASADEM and ASTER GDEM V3---A Case Study of the City of Dalian. Technology Innovation and Application, 11, 1-7+13.

Warren, S.D., Hohmann, M.G., Auerswald, K. and Mitasova, H. (2004) An evaluation of methods to determine slope using digital elevation data. CATENA, 58, 215–233.

Wei X., He H., Qu C. and Li. (2011) Geological environment feature and geologicical hazard in Yulin area of Shaanxi province. In: 2011 International Conference on Electric Technology and Civil Engineering (ICETCE). Presented at the 2011 International Conference on Electric Technology and Civil Engineering (ICETCE), IEEE, Lushan, China, pp. 1287–1291.

Xie, W., Li, X., Jian, W., Yang, Y., Liu, H., Robledo, L.F. and Nie, W. (2021) A Novel Hybrid Method for Landslide Susceptibility Mapping-Based GeoDetector and Machine Learning Cluster: A Case of Xiaojin County, China. ISPRS International Journal of Geo-Information, 10, 93.

Xiong, L., Li, S., Tang, G. and Strobl, J. (2022) Geomorphometry and terrain analysis: data, methods, platforms and applications. Earth-Science Reviews, 233:104191. https://doi.org/10.1016/j.earscirev.2022.104191.

Yalcin, A. and Bulut, F. (2007) Landslide susceptibility mapping using GIS and digital photogrammetric techniques: a case study from Ardesen (NE-Turkey). Nat Hazards, 41, 201–226.

Yan, G., Liang, S., Gui, X., Xie, Y. and Zhao, H. (2019) Optimizing landslide susceptibility mapping in the Kongtong District, NW China: comparing the subdivision criteria of factors. Geocarto International, 34, 1408–1426.

Yan, G., Tang, G., Li, S., Lu, D., Xiong, L. and Liang, S. (2023) Uncertainty in regional scale assessment of landslide susceptibility using various resolutions. Nat Hazards, 117, 399–423.

Yang, S., Yan, H. and Guo, Y. (2004) The Land Use Change and Its Eco-environmental Effects in Transitional Agro-pastoral Region -A Case Study of Yulin City in Northern Shaanxi Province. Progress in Geography, 49-55.

Zhou, X., Wen, H., Zhang, Y., Xu, J. and Zhang, W. (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geoscience Frontiers, 12, 101211.

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

Zhuang, J., Peng, J., Wang, G., Javed, I., Wang, Y. and Li, W. (2018) Distribution and characteristics of landslide in Loess Plateau: A case study in Shaanxi province. Engineering Geology, Special Issue: Loess Engineering Properties and Loess geohazards, 236, 89–96.