land

MDPI

Article

Machine Learning-Driven Landslide Susceptibility Mapping in the Himalayan China–Pakistan Economic Corridor Region

Mohib Ullah [1], Bingzhe Tang [1,*], Wenchao Huangfu [1], Dongdong Yang [1], Yingdong Wei [1] and Haijun Qiu [1,2]

[1] Shaanxi Key Laboratory of Earth Surface and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China; 2022880007@stumail.nwu.edu.cn (M.U.); 202110256@stumail.nwu.edu.cn (W.H.); yangdongdong@stumail.nwu.edu.cn (D.Y.); yingdongwei@stumail.nwu.edu.cn (Y.W.); haijunqiu@nwu.edu.cn (H.Q.)

[2] Institute of Earth Surface System and Hazards, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China

* Correspondence: tangbz121@nwu.edu.cn

Abstract: The reliability of data-driven approaches in generating landslide susceptibility maps depends on data quality, analytical method selection, and sampling techniques. Selecting optimal datasets and determining the most effective analytical methods pose significant challenges. This study assesses the performance of seven machine learning classifiers in the Himalayan region of the China–Pakistan Economic Corridor, utilizing statistical techniques and validation metrics. Thirteen geo-environmental variables were analyzed, including topographic (8), land cover (1), hydrological (1), geological (2), and meteorological (1) factors. These variables were evaluated for multicollinearity, feature importance, and their influence on landslide incidences. Our findings indicate that Support Vector Machines and Logistic Regression were highly effective, particularly near fault zones and roads, due to their effectiveness in handling complex, non-linear terrain interactions. Conversely, Random Forest and Logistic Regression demonstrated variability in their results. Each model distinctly identified landslide susceptibility zones ranging from very low to very high risk. Significant conditioning variables such as elevation, rainfall, lithology, slope, and land use were identified, reflecting the unique geomorphological conditions of the Himalayas. Further analysis using the Variance Inflation Factor and Pearson correlation coefficient showed minimal multicollinearity among the variables. Moreover, evaluations of Area Under the Receiver Operating Characteristic Curve (AUC-ROC) values confirmed the strong predictive capabilities of the models, with the Random Forest Classifier performing exceptionally well, achieving an AUC of 0.96 and an F-Score of 0.86. This study shows the importance of model selection based on dataset characteristics to enhance decision-making and strategy effectiveness.

Keywords: landslides; machine learning models; random forest classifier; landslide susceptibility

check for updates

Citation: Ullah, M.; Tang, B.; Huangfu, W.; Yang, D.; Wei, Y.; Qiu, H. Machine Learning-Driven Landslide Susceptibility Mapping in the Himalayan China–Pakistan Economic Corridor Region. *Land* **2024**, *13*, 1011. https://doi.org/10.3390/land13071011

Academic Editors: Faccini Francesco, Bouchra Haddad and Irene Manzella

Received: 3 May 2024
Revised: 8 June 2024
Accepted: 3 July 2024
Published: 8 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Landslides pose significant global challenges, resulting in tragic loss of life and substantial economic losses [1]. It ranks as the third most prevalent natural disaster, leading to economic damages exceeding USD 2.7 billion and impacting more than 3 million people globally, with 10,338 fatalities recorded from 2008 to 2017 [2]. The Abbottabad district, situated in the foothills of Balakot—devastated by the 2005 Kashmir earthquake—serves as a prime example of a region facing significant landslide risks within the Himalayan region in Pakistan. This susceptibility arises due to the rugged terrain, heavy rainfall during the monsoon season, high seismic activity, steep slopes, and unstable soil cover [3]. To mitigate these devastating effects, Landslide Susceptibility Mapping (LSM) has become a crucial tool. LSM allows experts to identify areas at high risk by evaluating terrain characteristics, thus playing a pivotal role in managing landslide risks [4]. This approach is particularly valuable in generating susceptibility maps that highlight areas prone to landslides within

specific regions [5]. Notably, regions with hilly terrain adjacent to large mountain ranges, such as the Himalayas, are particularly susceptible. The China–Pakistan Economic Corridor (CPEC) route traverses these mountains and has historically suffered from significant landslide incidents. Neighboring countries like India and Nepal have also faced severe landslide challenges, showing the urgent need for preventative strategies along the CPEC corridor [6]. Despite its importance, LSM research in these critical areas has faced several challenges, emphasizing the need for continued investigation in regions with a history of landslides.

The integration of diverse machine learning models marks a significant improvement in LSM. This advancement enhances the method's ability to reveal complex interactions between conditioning factors and landslide risks without primarily relying on traditional statistical assumptions [7,8]. The effectiveness of LSM depends not only on the choice of machine learning model but also on the careful selection of conditioning factors. These factors, which can range from a few to several hundred, are chosen based on the characteristics of the data and the unique attributes of the study area [9,10]. Common factors influencing landslide susceptibility include geological, topographical, and land cover elements [11]. The selection process of these factors is crucial—it involves assessing the importance of each factor and removing those that are unnecessary, thereby streamlining LSM and improving its accuracy [12]. This process often includes correlation analyses among factors to identify those most critical for predicting landslide susceptibility [13].

Extensive research has highlighted the crucial role of conditioning factors in LSMs, with many studies utilizing machine learning algorithms to assess landslide susceptibility in diverse geographical areas. Notably, Micheletti et al. applied Support Vector Machines and Random Forest algorithms in Switzerland's Vaud canton [14], while Lee, Hong, and Jung integrated Support Vector Machines with Geographic Information System for producing and evaluating landslide maps in PyeongChang and Inje, Korea [15]. Similarly, Meena et al. investigated the use of statistical and machine learning models in Italy's Belluno province [16]. In addition, various machine learning models have undergone rigorous testing for Landslide susceptibility modeling; for example, in northern Pakistan, a comparative study of Support Vector Machine, Random Forest, Maximum Entropy, Gradient Boosting Machines, and Logistic Regression was conducted, with optimization achieved through cross-validation techniques [17]. The Random Forest Classifier was also critically evaluated in Saudi Arabia and China [18,19]. Despite the widespread application of machine learning in landslide susceptibility analysis, there remains a gap in understanding these dynamics within the Himalayas, particularly in the CPEC. Our study addresses this gap by applying advanced machine learning techniques to investigate the unique interactions of conditioning factors and landslides in this region, aiming to enhance the accuracy of region-specific susceptibility models.

This research, building upon prior studies [17,20], continues to explore the most effective machine learning techniques for landslide susceptibility analysis in the Himalayan terrains of the CPEC. We employ seven machine learning models, including Support Vector Machines and Random Forest Classifiers, to analyze terrain vulnerabilities in the study area. Our study uniquely evaluates 13 conditioning factors specific to this region, such as slope, aspect, lithology, curvature, and land use, etc., thereby enhancing the accuracy of the landslide susceptibility models. Feature selection algorithms like Classification and Regression Trees (CART) and Recursive Feature Elimination (RFE) optimize the influential factors, supported by Variance Inflation Factor and correlation analysis. Additionally, the Wilcoxon test statistically assesses the model's significance. The goal is to improve hazard mitigation and land-use planning through comprehensive landslide susceptibility insights.

2. Materials and Methods

2.1. Study Area

The study area is situated in the western Himalayas, specifically within the China–Pakistan Economic Corridor (CPEC), as shown in Figure 1. It spans from 32.9° N to 39.7° N in latitude and 71.2° E to 77.8° E in longitude, characterized by towering mountains, glaciers, rivers, fault lines, and forests. This region is historically susceptible to natural disasters like earthquakes, landslides, and floods, particularly the 2005 earthquake, which had a magnitude of Mw = 7.6 and caused extensive damage to infrastructure, destabilizing numerous slopes in the area [21,22].

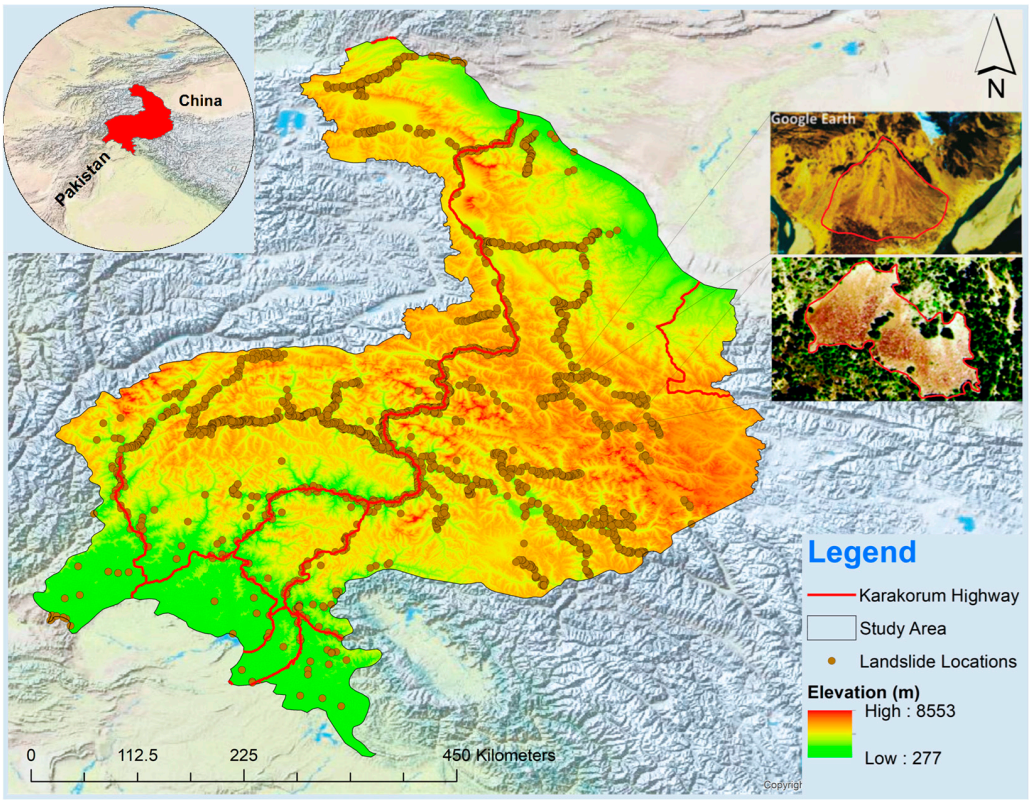


Figure 1. Study area and elevation map with highlighted landslide locations and major roads.

Geologically, the region is divided into two seismic zones: the Harman Seismic Zone (HSZ) and the Indus Kohistan Seismic Zone (IKSZ) [23]. Prominent fault lines, including the Main Mantle Thrust (MMT) and the Main Karakoram Thrust (MKT), play a crucial role in shaping the region's geological characteristics [24]. This region holds strategic importance within the upper Indus basin, notably due to the presence of mega infrastructures like the Diamer-Bhasha Dam and Dasu Hydropower Project [25]. The Karakoram Highway (KKH), a key component of the China–Pakistan Economic Corridor (CPEC), traverses the Karakoram, Hindu-Kush, and Himalayan mountain ranges, shows the region's importance in the northern sector of CPEC. Furthermore, the study area serves as the source of numerous glaciers and river systems, including the origin point of the Indus River. The climate is semi-arid, with southern areas receiving moisture from monsoon currents originating in the Indian Ocean, while the northern regions remain relatively arid [26]. This unique interplay of climatic zones, strategic importance, and varying altitudes makes it an essential area for geological and climatic hazard research.

2.2. Datasets

A landslide inventory crucial for landslide susceptibility mapping utilizes historical data, satellite imagery, and field surveys to examine the relationship between landslide occurrences and their causative factors. In this research, landslide locations were identified and mapped using high-resolution Sentinel-2 satellite (10 m) and Google Earth imagery [27,28], which are well-suited for detecting small to large-scale slope failures and landform alterations. The landslide inventory was validated using literature [29–33], historical data from the NASA geohazard catalog, and a comprehensive field survey of the study area, as shown in Figure 2 The inventory consists of 2197 polygon shapes drawn on satellite images based on validation data, each representing a landslide within the study area. The majority of landslides, accounting for 71.13%, are classified as very small, each covering an area of less than 1×10^5 m^2. Small landslides, which cover areas ranging from 1×10^5 to 3×10^5 m^2, make up 16.42% of the total landslides. Medium-sized landslides, covering areas between 3×10^5 and 5×10^5 m^2, constitute 10.72% of all landslides. Large landslides, each with an area spanning 5×10^5 to 1×10^6 m^2, are relatively uncommon, accounting for only 1.52%. The least frequent category is very large landslides, which cover an area greater than 1×10^6 m^2 and represent just 0.21% of all documented landslides. For the analysis, non-landslide areas were marked using ESRI ArcGIS, employing a random selection technique based on Yao et al.'s methodology [34]. This delineation process divided the locations into two categories: 1538 landslide locations for model training (70%) and 659 for validation (30%).

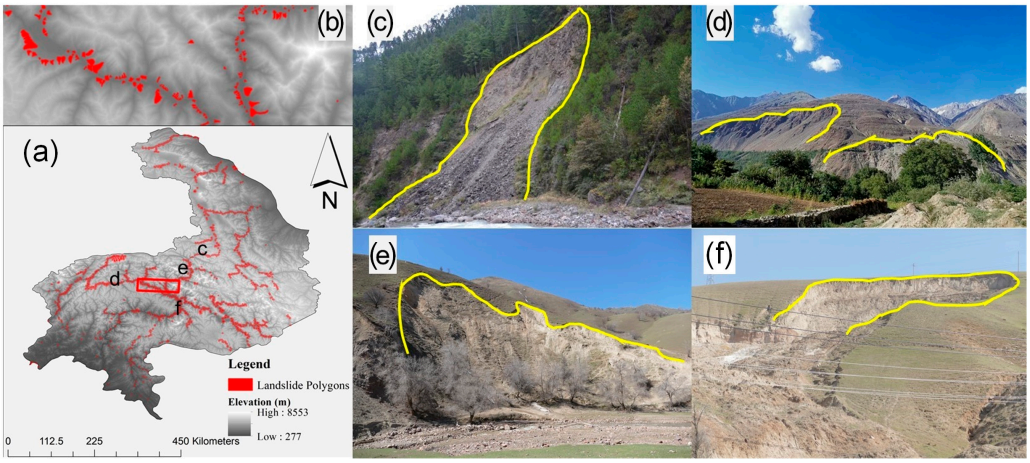


Figure 2. Field investigation maps depicting landslides in the study area, including (**a**) Study area map showing landslide polygons in red and elevation in grayscale. The red square indicates the specific region detailed in subfigures (**c**–**f**). (**b**) Detailed topographic map of the area within the red square in (**a**), showing the distribution of landslide polygons marked in red. The yellow lines in (**c**–**f**) delineate the boundaries of observed landslides. (**c**) Bara Khun area, (**d**) Shandur area, (**e**) Morkhun area, and (**f**) Astore area.

For landslide susceptibility analysis in this study, thirteen conditioning factors were carefully selected, as shown in Table 1 and Figures 3 and 4, which are categorized into topographic, land cover, hydrological, geological, and meteorological types. These factors include valley depth, terrain relief, diverse land use classifications, and the impact of rainfall, proximity to road networks, lithology, and active fault lines [35–37]. Analytical indices such as the Terrain Ruggedness Index (TRI) and the Topographic Wetness Index (TWI), alongside factors like elevation variations, land use changes, and slope gradients, were also crucial [38,39]. This comprehensive approach provides a robust framework for assessing landslide risks, ensuring a detailed analysis of susceptibility in the region.

Table 1. Shows conditioning factors with sources and spatial resolutions used for landslide susceptibility in the study.

Sl. No.	Conditioning Factors	Source	Spatial Resolution
1	Slope	SRTM 30 m DEM	30 mt.
2	Aspect	SRTM 30 m DEM	30 mt.
3	Curvature	SRTM 30 m DEM	30 mt.
4	Land use land cover	USGS Global Land Cover (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-global-land-cover) (accessed on 2 February 2024)	30 mt. (resampled)
5	Lithology	National Library of Australia (http://nla.gov.au/nla.obj-233602228) (accessed on 2 February 2024)	30 mt. (resampled)
6	Distance from fault	National Library of Australia (http://nla.gov.au/nla.obj-233602228) (accessed on 2 February 2024)	30 mt. (resampled)
7	Distance from the road	Natural Earth Data (https://www.naturalearthdata.com/downloads/10m-cultural-vectors/) (accessed on 2 February 2024)	30 mt. (resampled)
8	Terrain Relief	SRTM 30 m DEM	30 mt.
9	Valley Depth	SRTM 30 m DEM	30 mt.
10	Topographic Wetness Index (TWI)	SRTM 30 m DEM	30 mt.
11	Terrain Ruggedness Index (TRI)	SRTM 30 m DEM	30 mt.
12	Elevation	SRTM 30 m DEM	30 mt.
13	Rainfall	Climate Hazards Group InfraRed Precipitation with Station data	30 mt. (rescaled)

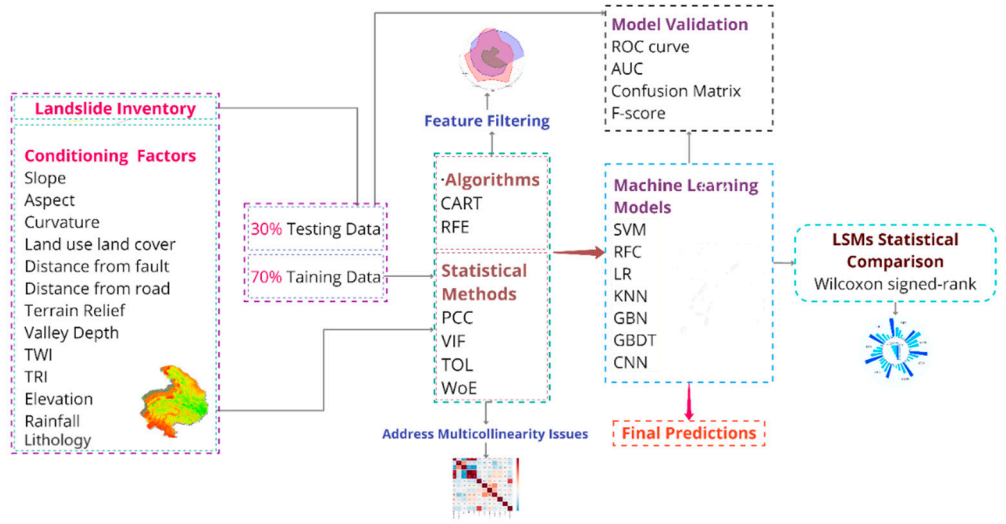


Figure 3. Workflow diagram illustrating the process of landslide susceptibility mapping. CART: Classification and Regression Trees, RFE: Recursive Feature Elimination, PCC: Pearson Correlation Coefficient, VIF: Variance Inflation Factor, TOL: Tolerance, WoE: Weight of Evidence, ROC curve: Receiver Operating Characteristic curve, AUC: Area Under the Curve, SVM: Support Vector Machine, RFC: Random Forest Classifier, LR: Logistic Regression, KNN: K-Nearest Neighbors, GBN: Gaussian Naive Bayes, GBDT: Gradient Boosting Decision Tree, CNN: Convolutional Neural Network.

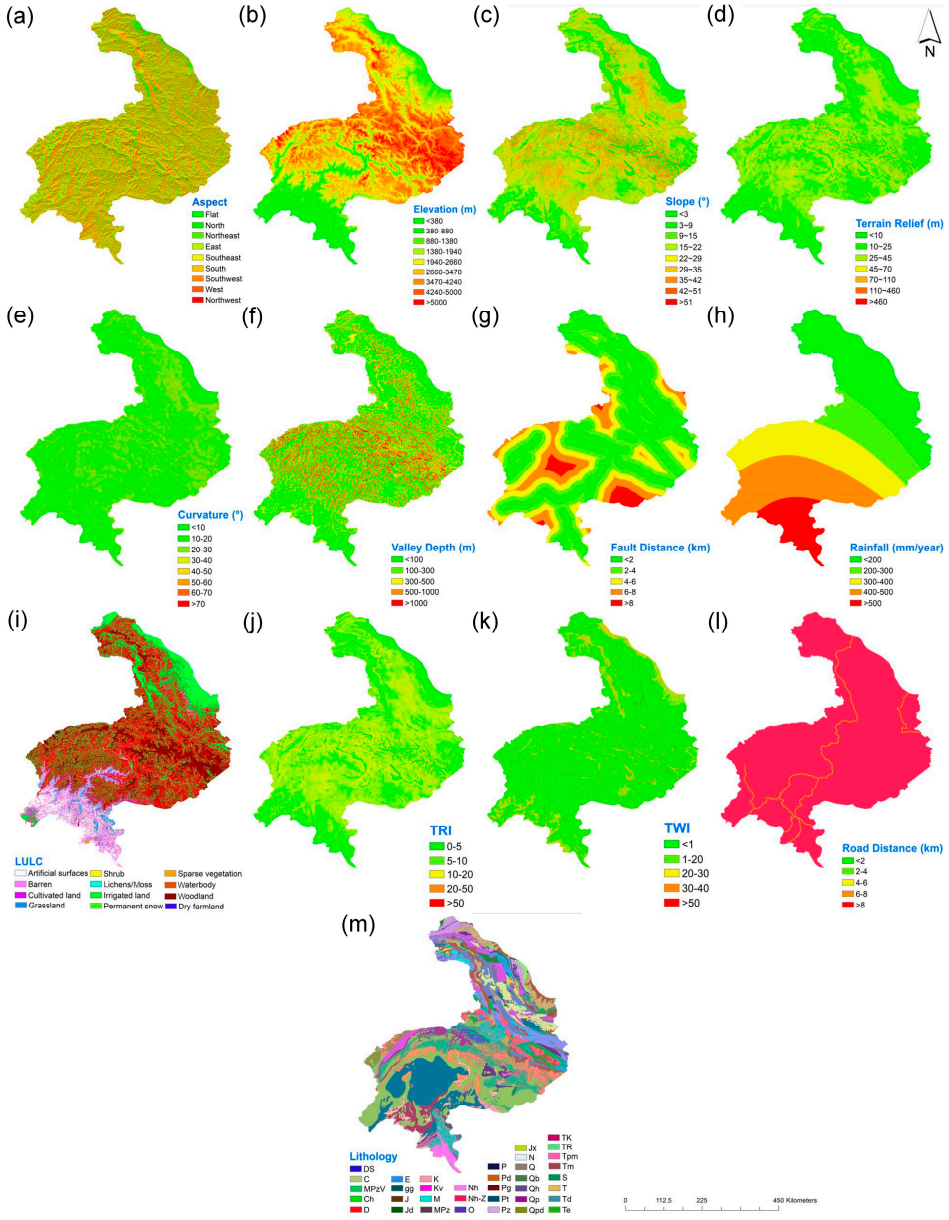


Figure 4. Maps of conditioning factors in the study area: (**a**) aspect, (**b**) elevation, (**c**) slope, (**d**) terrain relief, (**e**) curvature, (**f**) valley depth, (**g**) active fault zones, (**h**) rainfall, (**i**) Land Use Land Cover (LULC), (**j**) Terrain Ruggedness Index (TRI), (**k**) Topographic Wetness Index (TWI), (**l**) road distance, (**m**) lithology. Additional Notes for (**m**) DS—Devonian Schist, C—Carboniferous Limestone, MPzV—Paleozoic Slate, Ch—Metamorphic Schist, E—Eocene Marl, D—Devonian Sandstone, gg—Igneous Metamorphic, J—Jurassic Limestone, Jd—Jurassic Dolomite, K—Cretaceous Sandstone, Kv—Cretaceous Volcanic, M—Mesozoic Sandstone, MPz—Paleozoic Sedimentary, Nh—Neogene Conglomerate, Nh-Z—Neogene Marble, O—Ordovician Limestone, P—Permian Limestone, Pd—Permian Dolostone, Pg—Permian Siliceous, Pt—Permian Quartzite, Pz—Paleozoic Metamorphic, Jx—Jurassic Metacarbonate, N—Neogene Sandstone, Q—Quaternary Sandstone, Qb—Quaternary Carbonate, Qh—Quaternary Mudstone, Qp—Pleistocene Gravel, Qpd—Quaternary Clastic, TK—Tertiary Cretaceous, TR—Triassic Sandstone, Tpm—Pliocene Calcareous, Tm—Miocene Marl, S—Silurian Limestone, T—Tertiary Carbonate, Td—Tertiary Dolostone, Te—Eocene Limestone.

The Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), available through NASA's Earthdata Search website (https://earthdata.nasa.gov/) (accessed on 2 February 2024), offers a spatial resolution of 30 m. These data were used in deriving several critical factors for landslide susceptibility analysis, including slope, aspect, curvature,

terrain relief, valley depth, the Topographic Wetness Index (TWI), the Terrain Ruggedness Index (TRI), and elevation. Further parameters were incorporated using resampled 30 m resolution data: these data sources include USGS Global Land Cover, the National Library of Australia for Lithology and Fault Distance, Natural Earth Data for Road Distance, and the Climate Hazards Group for Rainfall data.

Thematic maps for the selected conditioning factors were produced using ArcGIS software 10.8.1., with all data resampled to a consistent 30 m resolution for uniformity. The datasets were standardized and normalized to minimize redundancy and organize the data effectively. Each raster layer representing a conditioning factor was categorized into various levels, with all maps created for this study following the WGS 1984 datum and projected in the UTM zone 42 system.

2.3. Statistical Feature Filtering and Assessment

Multicollinearity in the dataset, arising from interdependent conditioning factors such as slope, elevation, and rainfall, was addressed to enhance the accuracy of landslide susceptibility modeling.

To analyze the relationships between these variables, the Pearson Correlation Coefficient (PCC) was employed, which quantifies the linear correlation between pairs of variables, providing insights into the strength and direction of their relationships [40]. The PCC is defined by the following formula [41]:

$$PCC = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\sum_{j=1}^{n}(y_j - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{j=1}^{n}(y_j - \overline{y})^2}} \quad (1)$$

where x_i and y_j are individual sample values and $\overline{x}$ and $\overline{y}$ their respective means.

Tolerance (TOL) and the Variance Inflation Factor (VIF) were used for multicollinearity assessment, which is crucial in identifying and mitigating the influence of strongly correlated variables to ensure the robustness and reliability of landslide susceptibility models [42]. The computation for VIF is as follows [43]:

$$VIF_i = \frac{1}{1 - R_i^2} = \frac{1}{TOL}(i = 1, 2, 3 \ldots\ldots k) \quad (2)$$

where R_i^2 is the determination coefficient for the ith variable.

Moreover, the Weight of Evidence (WoE) approach, which is based on Bayesian probability principles, investigates spatial relationships between landslides and key factors using a logarithmic linear framework. It calculates positive (α) and negative (β) correlations, and their combined effect (Δ) on landslide occurrences as follows [44]:

$$\alpha = \ln\left[\frac{Q(X|Y)}{Q(X|Y')}\right] \quad (3)$$

$$\beta = \ln\left[\frac{Q(X'|Y)}{Q(X|Y')}\right] \quad (4)$$

$$\Delta = \alpha + \beta \quad (5)$$

where Q represents probability, X signifies specific factor categories, X' represents categories not within the factors, Y relates to landslide-prone regions, and Y' refers to areas without landslides.

Furthermore, a combination of advanced algorithms was employed for optimized feature selection and statistical comparison of machine learning models. Classification and Regression Trees (CART) were used to construct binary decision trees and identify critical variables, while Recursive Feature Elimination (RFE) was implemented for its systematic approach to iteratively refining the feature set [45]. Furthermore, for the statistical comparison of different landslide susceptibility models, the Wilcoxon signed-rank test was

employed. This non-parametric method assesses the significance of differences between models at a set level (typically $\alpha = 0.05$), using *p*-values and z-values to determine whether the null hypothesis of no significant difference can be rejected or affirmed [46]. This comprehensive approach, combining feature selection algorithms with rigorous statistical testing, provided a robust framework for landslide susceptibility analysis.

2.4. Machine Learning Techniques

In this study, seven well-established machine learning models were used, each selected for its specific strengths in predicting landslide susceptibility. These classifiers, comprise Support Vector Machine, Logistic Regression, Random Forest, K-Nearest Neighbors, Gradient Boosting Decision Tree, Gaussian Bayesian Network, and Convolutional Neural Network, which are briefly outlined in terms of their fundamental principles as follows:

The Support Vector Machine, utilizing the Radial Basis Function (RBF) kernel, is effective for high-dimensional data and is represented by the following equation [47,48]:

$$k(x_i, x_j) = exp\left(-g\left\|x_i - x_j\right\|^2\right) \tag{6}$$

where the kernel function $k(x_i, x_j)$ applies an exponential decay, exp, to the negative scaled squared distance $\left\|x_i - x_j\right\|^2$ between vectors x_i and x_j, modulated by g. Moreover, Logistic Regression, used for binary classification and thus effective in determining event occurrence, models landslide probabilities as follows [49]:

$$P = \frac{e^Y}{1 + e^Y} \tag{7}$$

$$Y = B + A_1X_1 + A_2X_2 + \cdots + A_nX_n \tag{8}$$

where P is the predicted landslide probability, Y is the logit function, B is the intercept, A_i are coefficients, and X_i are predictors affecting landslide risk.

The Random Forest Classifier (RFC) enhances accuracy through its ensemble approach, making it suitable for diverse data subsets [50,51]. It can be described by the following equation:

$$F(x) = \frac{1}{B}\sum_{b=1}^{B} \quad f_{b(x)} \tag{9}$$

where $F(x)$ is the averaged result of multiple predictions $f_{b(x)}$ made by an ensemble of B different models, with b as the index iterating through each model to compute the ensemble's output.

The Convolutional Neural Network features exceptional capabilities in feature extraction and resistance to overfitting, which makes it essential for processing input data in the following formula [52].

$$Rescaled \;\; x = \frac{x - x\text{min}}{x\text{max} - x\text{min}} \tag{10}$$

Here, $Rescaled \;\; x$ normalizes x between 0 and 1, where $xmin$ and $xmax$ are its minimum and maximum values, shifting the range so that all values are proportionally adjusted.

The Gradient Boosting Decision Tree (GBDT) iteratively improves accuracy, making it efficient for modeling complex relationships using the following equation [53,54]:

$$Y_t(X_i) = Y_{t-1}(X_i) + \eta F_t(X_i) \tag{11}$$

This indicates the prediction $Y_t(X_i)$ refined through prior predictions $Y_{t-1}(X_i)$ and the current weak learner's contribution $F_t(X_i)$, with η moderating the update.

Furthermore, GBDT employs the F-MSE impurity metric [55]:

$$F - MSE = \frac{w_L w_R}{w_L + w_R} \cdot \left(\frac{\sum_L (r_i - \hat{y}_i)^2}{w_L} - \frac{\sum_R (r_i - \hat{y}_i)^2}{w_R}\right)^2 \tag{12}$$

Here, in the GBDT $F - MSE$ formula, w_L and w_R are weights for the left and right nodes, respectively. They normalize the sums $\sum L$ and $\sum R$ of squared errors $(r_i - \hat{y}_i)^2$ to calculate node-specific mean squared errors. The final $F - MSE$, a squared, weighted difference of these errors, emphasizes efficient branching to improve prediction accuracy.

The cubic K-Nearest Neighbor model is advantageous for geological data analysis, facilitating precise classification as follows [56,57]:

$$Distance = \sum_{i=1}^{n} |u_i - v_i|^3 \tag{13}$$

In this formula, u_i and v_i are the i-th components of vectors u and v, respectively.

Finally, the Gaussian Naive Bayes model, known for its straightforward and effective application, determines landslide likelihood using the following equation [58]:

$$z_{\mathrm{NB}} = \underset{w_j \in \{\text{landslide presence,absence}\}}{argmax} P_{wj} \times \prod_{i=1}^{n} P_{\frac{z_i}{w_j}} \tag{14}$$

Here, the Naive Bayes classifier z_{NB} predicts the class label (either "landslide presence" or "absence") by selecting the class wj that maximizes the combined probabilities of the prior P_{wj} and the conditional probabilities $\prod_{i=1}^{n} P_{\frac{z_i}{w_j}}$ for each feature zi. This method uses $argmax$ predict the most likely class, assuming features are independent. Moreover, the likelihood of the predictors, given the outcome category, is expressed as a Gaussian distribution [59]:

$$P_{\frac{z_i}{w_j}} = \frac{1}{\sqrt{2\pi\rho^2}} \times e^{-\frac{z_i - \lambda^2}{2\rho^2}} \tag{15}$$

In this formula, λ represents the mean, and ρ represents the standard deviation associated with each feature zi for the outcome wj.

2.5. Validation and Reliability Evaluation

The validation of landslide susceptibility models was conducted using two key indicators. The Receiver Operating Characteristic (ROC) curve was employed to measure classification effectiveness, with the Area Under the Curve (AUC) indicating accuracy levels [60]. A higher AUC value, approaching 1, denotes greater model accuracy, while values near or below 0.5 suggest limited predictive ability. Additionally, the F-Score is a harmonic mean of precision and recall, which ranges from 0 to 1, with values closer to 1 indicating better performance [61]. For the validation of models, these indicators were applied to 30% testing data.

3. Results

3.1. Assessment of Explanatory Variables

The relationships among thirteen geographical and environmental factors were analyzed using Pearson correlation coefficients, as illustrated in Figure 5. A key finding was the strong positive correlation (0.92) between slope and terrain relief, indicating that steeper slopes are often found in areas of significant relief. Conversely, a negative correlation (-0.55) between slope and the Topographic Wetness Index (TWI) suggested that higher slope areas tend to have lower TWI values. Some factors display significant correlations with each other's, while others exhibit minimal associations; this dynamic interrelation among conditioning factors emphasizes the diverse influences on landslide susceptibility.

Furthermore, the details of multicollinearity assessment are presented in Table 2. All conditioning variables showed minimal multicollinearity (VIF values below 3), indicating that each factor contributes uniquely to the models. Elevation displayed the highest VIF value of 2.98, indicating significant interactions with other variables, suggesting a central role in the analysis due to its low tolerance of 0.33. Conversely, 'Distance to active_faults', with its lower VIF of 1.10 and higher TOL of 0.96, exhibited a relatively independent

influence within the model. This highlights the importance of selecting relevant variables for accurately modeling landslide susceptibility.

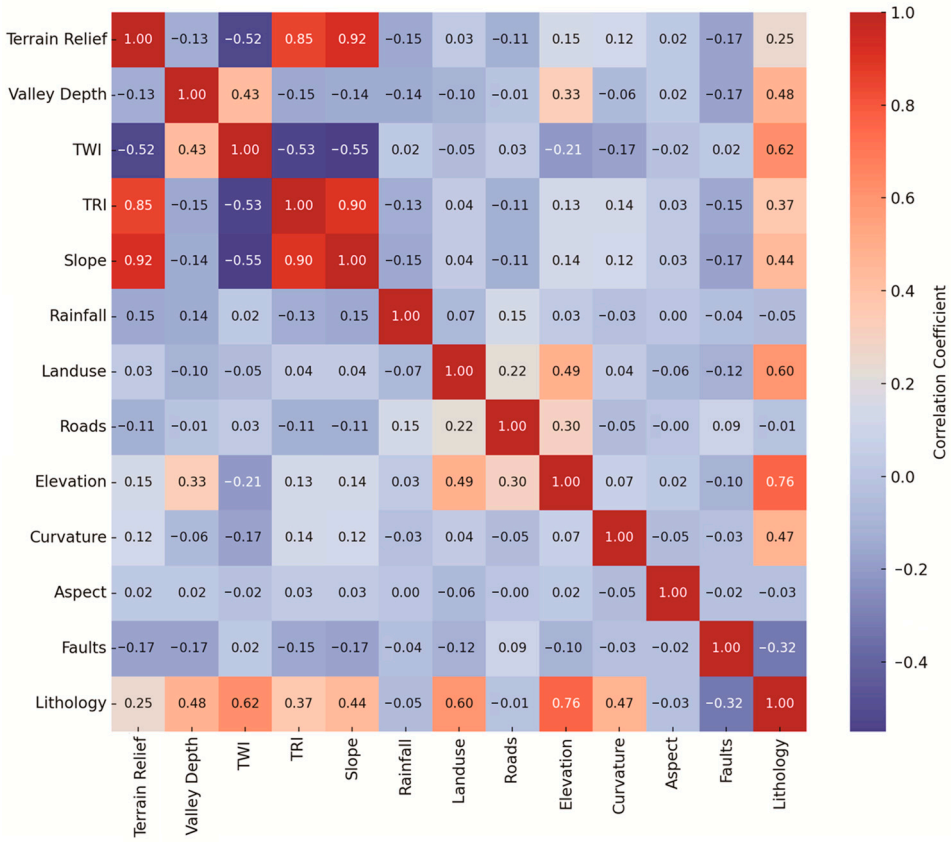


Figure 5. Heat map of Pearson correlations for landslide variables, with red-to-blue grids showing high-to-low strength and labels representing r-values from Pearson correlations analysis, ranging from −1 to +1.

Table 2. Variance inflation factor and tolerance values for landslide susceptibility factors.

Variable	VIF	TOL
Distance to active faults	1.10	0.96
Valley Depth	1.18	0.85
TWI	1.36	0.73
Distance to Roads	1.55	0.65
Curvature	1.73	0.58
Aspect	1.91	0.52
Terrain Relief	2.09	0.48
Land Use	2.27	0.44
Slope	2.46	0.41
TRI	2.64	0.38
Lithology	2.74	0.36
Rainfall	2.82	0.36
Elevation	2.98	0.33

3.2. Factors Influencing Landslide Susceptibility

The impact of various causative factors on landslides across different categories can be observed in Figure 6. Proximity to active fault zones and road construction within 2 km leads to landslide occurrences of 12.4% and 12.1%, respectively, due to geological instability and altered terrain. Anthropogenic activities within elevation ranges between 1940 and 2660 m contribute to an 18.2% landslide incidence. Additionally, sub-classes such

as shrub, sparse vegetation, and snow increase susceptibility to landslides by reducing soil stability, limiting root reinforcement, and promoting saturation, respectively. Aspect direction also plays a role, with south and west subclasses exhibiting landslide occurrences of 20.3% and 12.8%, respectively. This can be attributed to increased exposure to solar radiation and weathering on south-facing slopes, as well as enhanced susceptibility to rainfall-induced erosion and soil saturation on west-facing slopes, influenced by prevailing wind and precipitation patterns. In TWI > 50 sub-class areas, landslide occurrence is notably high at 61.9% due to increased saturation and reduced soil stability linked to higher topographic wetness indices, amplifying slope instability. Valley depths exceeding 1000 m experience a high landslide occurrence rate of 20.4%. Geological formations such as Eocene Limestone exhibit high landslide occurrence rates due to weathering and erosion processes, while human activities like quarrying and construction further amplify landslide risks in limestone-rich areas.

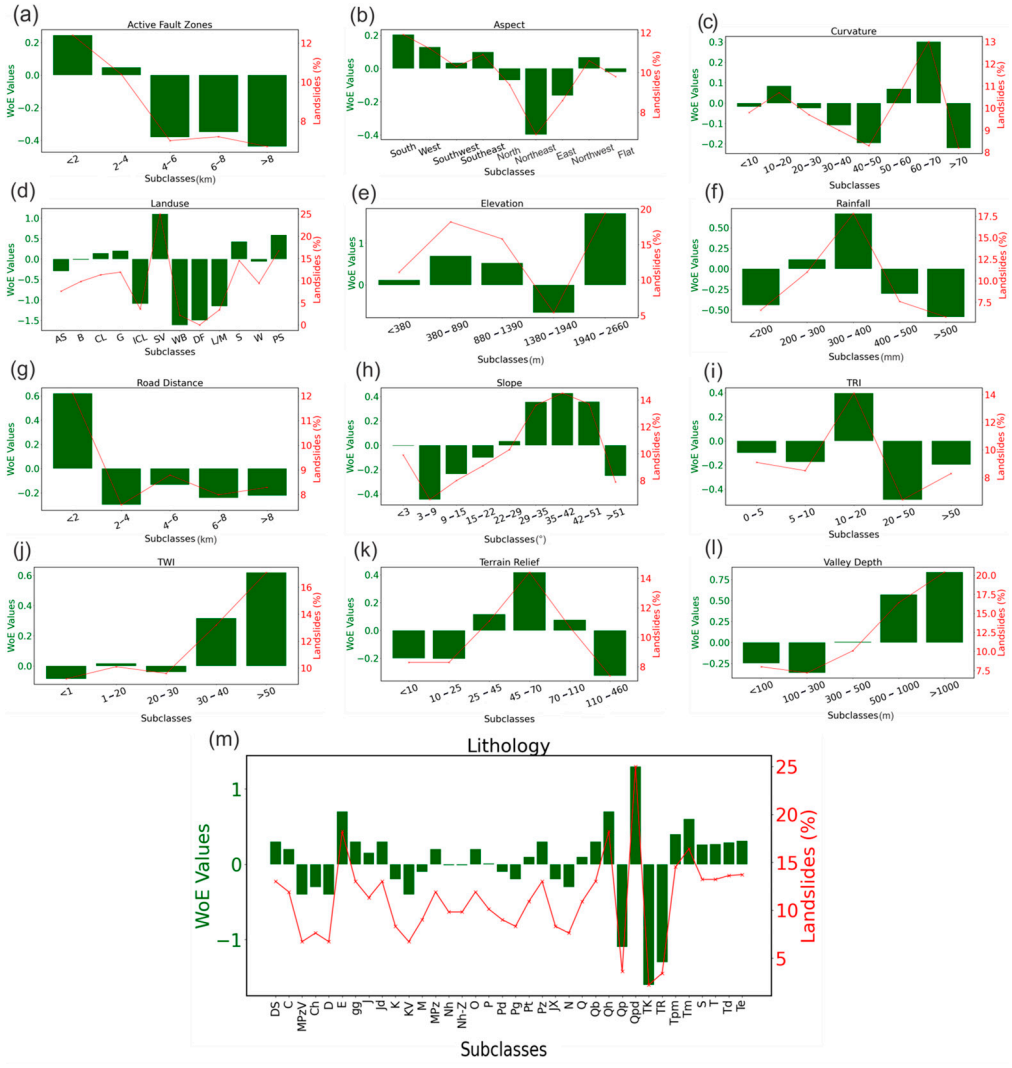


Figure 6. Illustrating factor–landslide relationships through the weight of evidence model. Each subplot is labeled with the corresponding factor name, including (**a**) active fault zone, (**b**) aspect, (**c**) curvature, (**d**) land use, (**e**) elevation, (**f**) rainfall, (**g**) road distance, (**h**) slope, (**i**) TRI (Topographic Ruggedness Index), (**j**) TWI (Topographic Wetness Index), (**k**) terrain relief, (**l**) valley depth, and (**m**) lithology. Land use classes are represented with their abbreviations: Artificial surfaces (AS), Barren (B), Cultivated Land (CL), Grassland (GL), Irrigated Cultivated Land (ICL), Sparse Vegetation (SV), Waterbody (WB), Dry Farmland (DFL/M), Lichens/Moss (LM), Shrub (S), Woodland (W), and Permanent Snow (PS).

3.3. Feature Priority Selection

The key factors for landslide susceptibility mapping were selected using the Classification and Regression Trees (CART) Importance and Recursive Feature Elimination (RFE) Ranking methods, as depicted in Figure 7. According to the CART Importance analysis, Rainfall emerged as the most critical factor with a value of 0.16, followed by elevation at 0.13, slope at 0.11, lithology at 0.09, and land use at 0.10. Conversely, the RFE Ranking method identified elevation as the most important feature, assigning it a rank of 1. It was closely followed by rainfall with a rank of 2, slope at 3, lithology at 4, and land use at 5. Valley depth was evaluated as the least significant factor, with a CART Importance value of 0.02 and an RFE Ranking of 13. This analysis demonstrates that while certain factors vary in importance, each one contributes to the predictive accuracy of landslide modeling. Therefore, all factors were selected for further detailed analysis.

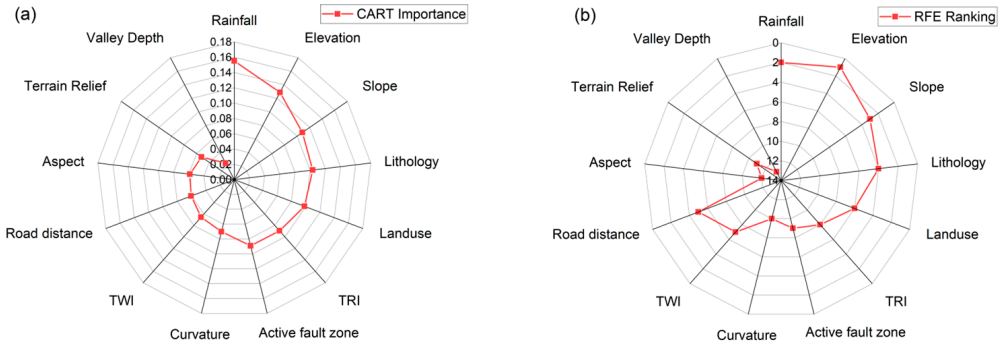


Figure 7. Landslide covariate priority selection by CART and RFE models. (**a**) displays the priority of different covariates as determined by the CART (Classification and Regression Trees) model, quantifying each covariate's importance in predicting landslide occurrence. The values on the radar chart reflect the normalized importance scores ranging from 0 to 0.18, with higher values indicating greater importance in the model. (**b**) illustrates the ranking of covariates by the RFE (Recursive Feature Elimination) method, where each covariate's contribution to model accuracy is ranked from 0 to 12, with lower numbers representing higher priority.

3.4. Landslide Susceptibility Analysis

In the process of landslide susceptibility mapping, conditioning factors are integrated into study area pixels, various models predict landslide probabilities, and the area is segmented into five susceptibility categories—very low to very high—using the quantile method. In this process, selected landslide-causing factors are input variables, and grid units are classified as 'landslide' (1) or 'non-landslide' (0) in the model's output. The result findings align closely with the geographical attributes of the study area. Predominantly, very-high susceptibility zones are concentrated in mountainous regions, whereas very-low susceptibility zones are covered in vegetated areas. Table 3 and Figure 8 show that the VERY LOW susceptibility category consistently covers a significant portion of the study area, ranging from 26.68% to 59.61%, with corresponding landslide frequencies varying between 0.8% and 6.76%. Moreover, at the LOW susceptibility level, the models show varying spatial extents, ranging from 12.68% to 41.62%, along with corresponding landslide frequencies of 2.9% to 7.46%. The MODERATE susceptibility zones cover a range of 7.61% and 15.89% of the total area, with corresponding landslide frequencies that vary from 4.42% to 11.87%. HIGH susceptibility areas exhibit substantial spatial variability between 6.22% and 11.05%, while the landslide frequencies also vary significantly, ranging from 10.81% to 20.98%. The VERY HIGH susceptibility zones consistently occupy smaller proportions of the study area, ranging from 6.10% to 13.59%, while showing the highest landslide frequencies, ranging from approximately 53.57% to 75.33%.

Table 3. Susceptibility classes and landslide percentages as identified by machine learning models in this study.

SVM			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	62,554,112	26.68	1.93
Low	97,580,722	41.62	6.9
Moderate	28,868,401	12.31	7.55
High	14,575,374	6.22	10.81
Very High	30,893,955	13.18	72.8
RFC			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	139,775,856	59.61	3.5
Low	43,713,122	18.64	3.73
Moderate	21,461,542	9.15	4.42
High	15,208,584	6.49	13.02
Very High	14,313,460	6.1	75.33
LR			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	93,261,404	39.77	6.63
Low	57,940,244	24.71	7.32
Moderate	37,267,200	15.89	11.5
High	25,901,038	11.05	20.98
Very High	20,102,678	8.57	53.57
KNN			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	139,571,787	59.53	4.28
Low	32,999,116	14.07	5.43
Moderate	20,573,972	8.77	7.55
High	15,290,956	6.52	11.04
Very High	26,036,733	11.1	71.7
GBN			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	138,211,363	58.95	6.76
Low	29,733,311	12.68	7.46
Moderate	17,832,691	7.61	11.87
High	16,821,114	7.17	12.43
Very High	31,874,085	13.59	61.48
GBDT			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	110,612,605	47.18	3.73
Low	55,933,064	23.85	5.66
Moderate	28,009,937	11.95	9.76
High	17,227,718	7.35	16.94
Very High	22,689,240	9.68	63.92
CNN			
Susceptibility	Pixel Amount	Area (%)	Landslide Frequency (%)
Very Low	136,521,785	58.23	0.8%
Low	41,280,794	17.61	2.9%
Moderate	21,444,077	9.15	8.2%
High	16,783,057	7.16	18.2%
Very High	18,442,851	7.87	67.4%

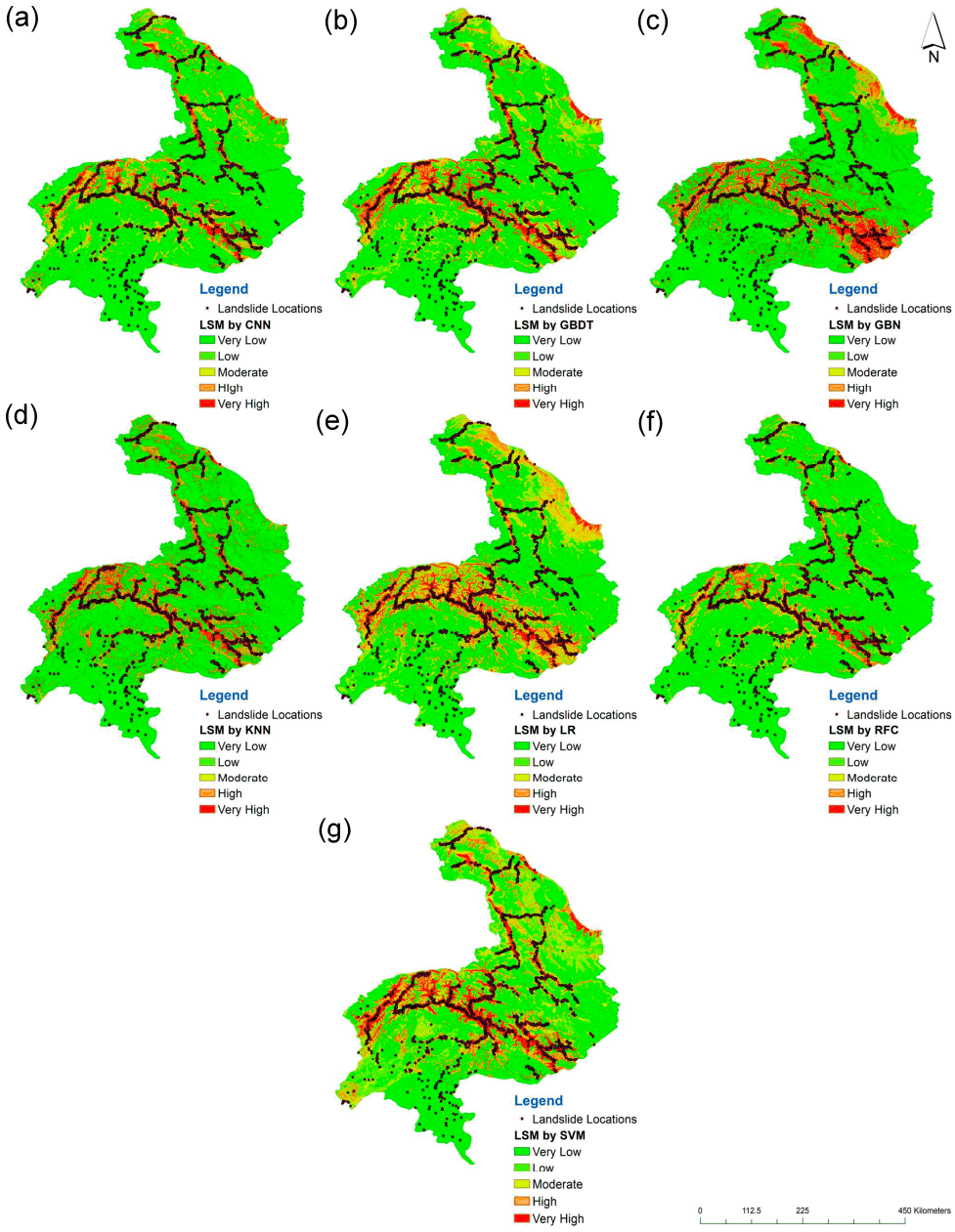


Figure 8. Landslide susceptibility maps obtained using each model selection: (**a**) CNN, (**b**) GBDT, (**c**) GBN, (**d**) KNN, (**e**) LR, (**f**) RFC, (**g**) SVM.

Furthermore, Random Forest Classifiers and Convolutional Neural Networks tend to categorize larger regions as very low risk, while Logistic Regression tends to identify smaller zones in this risk category. For low-risk areas, a Support Vector Machine covers a larger area compared to a Random Forest Classifier and K-Nearest Neighbors. The moderate-risk zone exhibits notable variability across the models. In the high-risk category, the Convolutional Neural Network and Gaussian Bayesian Network delineate larger areas with a heightened risk of landslides, while the Support Vector Machine designates a smaller extent. In the very high-risk zone, K-Nearest Neighbors and the Support Vector Machine allocate larger areas, with the Random Forest Classifier indicating a higher frequency of landslides, whereas the Random Forest Classifier and Gradient Boosting Decision Tree allocate smaller extents.

3.5. Models Correlations in LSM

This study classifies the pairwise correlations of machine learning model performance into three distinct categories—strong, moderate, and weak—based on the *p*-values presented in Table 4. Strong correlations were observed among model pairs such as Support Vector Machine vs. Logistic Regression, Support Vector Machine vs. Gaussian Bayesian Network, and Gaussian Bayesian Network vs. Gradient Boosting Decision Tree, with *p*-values approaching 1, which suggests similar methodologies and possibly common mechanisms of models in landslide susceptibility assessment. The high correlation indicates these models have significant overlap in data processing and interpretation, likely due to similar algorithmic structures or dataset feature sensitivities. Moderate correlations, with *p*-values between 0.25 and 0.75, were observed in pairs like Support Vector Machine vs. Random Forest Classifier and Logistic Regression vs. Gaussian Bayesian Network, which shows these models share certain methodological aspects but also have notable differences in their evaluations, indicating a mix of shared and unique approaches in their processing and analysis. Furthermore, weak correlations, seen in pairs like Random Forest Classifier vs. Logistic Regression and Gaussian Bayesian Network vs. Convolutional Neural Network with *p*-values close to 0, highlight distinct model approaches in landslide susceptibility, emphasizing the need for model selection that aligns with specific study goals for effective analysis.

Table 4. Wilcoxon signed-rank test used for comparing models in the study.

Model Pair	*p*-Value	Significance	Correlation
SVM vs. RFC	0.25	high	Moderate
SVM vs. LR	0.875	high	Strong
SVM vs. KNN	0.25	high	Moderate
SVM vs. GBN	0.875	high	Strong
SVM vs. GBDT	0.875	high	Strong
SVM vs. CNN	0.25	high	Moderate
RFC vs. LR	0.125	high	Weak
RFC vs. KNN	0.125	high	Weak
RFC vs. GBN	0.125	high	Weak
RFC vs. GBDT	0.125	high	Weak
RFC vs. CNN	0.125	high	Weak
LR vs. KNN	0.125	high	Weak
LR vs. GBN	0.625	high	Moderate
LR vs. GBDT	0.625	high	Moderate
LR vs. CNN	0.125	high	Weak
KNN vs. GBN	0.125	high	Weak
KNN vs. GBDT	0.125	high	Weak
KNN vs. CNN	0.125	high	Weak
GBN vs. GBDT	0.875	high	Strong
GBN vs. CNN	0.125	high	Weak
GBDT vs. CNN	0.125	high	Weak

3.6. Models Validation

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and F-score values obtained from the test data indicate an acceptable level of predictive performance, as detailed in Table 5. In the analysis of machine learning models for their effectiveness, the Convolutional Neural Network and Random Forest Classifier show superior performance with F-scores of 0.84 and 0.86, respectively, and AUCs of 0.93 and 0.96 Figure 9. These models outperformed Logistic Regression, which had a lower F-score of 0.57 and an AUC of 0.87, indicating its lesser efficacy in this context. The Support Vector Machine and Gaussian Naive Bayes models also demonstrated moderate effectiveness, with the Support Vector Machine achieving an F-score of 0.69 and an AUC of 0.91, while the Gaussian Bayesian Network lagged slightly with an F-score of 0.63 and an AUC of 0.86, reflecting their varying degrees of suitability for landslide susceptibility analysis.

Table 5. Evaluation metrics for machine learning models in landslide susceptibility assessment.

Model	AUC	F-Score
SVM	0.91	0.69
CNN	0.93	0.84
RFC	0.96	0.86
LR	0.87	0.57
KNN	0.92	0.74
GBN	0.86	0.63
GBDT	0.93	0.61

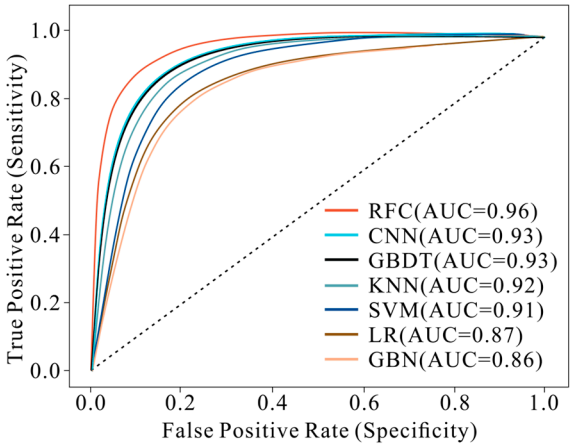


Figure 9. Receiver Operating Characteristic (ROC) curves for various machine learning models evaluating landslide susceptibility.

4. Discussion

The findings of this research contribute significantly to the understanding of landslide susceptibility within the Himalayan region enclosed by the China–Pakistan Economic Corridor. Data analysis showed the Random Forest Classifier as the leading model, outperforming other models like Support Vector Machine, K-Nearest Neighbors, and Gaussian Bayesian Network [62]. Random Forest Classifier's superior performance is attributed to its ensemble method, effectiveness in handling non-linear patterns, and ability to process high-dimensional data, making it particularly suitable for the Himalayan dataset. However, the effectiveness of these machine learning models is critically dependent on the quality of the input data. For landslide susceptibility mapping, it is essential to precisely capture the unique characteristics and dynamics of each landslide to ensure reliable analysis and effective mitigation strategies. Meanwhile, accurate representation of their extent and shape is crucial for evaluating impacted areas and developing effective mitigation measures. Additionally, not specifying the state of activity and the typology of landslides can result in risk assessments that are either underestimated or overestimated. To improve accuracy and reliability in landslide risk assessments, future studies will focus on monitoring various landslide types and incorporating detailed classifications of their activity states.

Furthermore, a strong correlation between slope and terrain relief reflects the area's unique geomorphology, arising from the tectonic collision of the Indian and Eurasian plates [63,64]. Steeper slopes correspond to higher relief, inherently increasing landslide susceptibility [65]. The minimal correlation between aspect and rainfall indicates their separate influences on landslide susceptibility. Aspect influences micro-environments by affecting sunlight exposure and vegetation types, impacting landslide risk [66,67]. For instance, south-facing slopes are generally warmer, supporting drought-resistant flora, while north-facing slopes are cooler and moister, suitable for shade-loving species [68]. On the other hand, in the Himalayas, the moderate correlation between valley depth and elevation suggests a complex link to landslide susceptibility. Higher elevations, with deeper valleys,

face increased landslide risks due to factors like soil saturation and reduced vegetation. This relationship, while significant, is part of a broader set of variables, including geology and human activities, highlighting the intricate dynamics of landslides in this region.

Landslide susceptibility in the study area is influenced by a complex interplay of environmental and geographical factors, such as sparse vegetation, notably in shrub areas, which increases landslide risk by reducing soil stabilization. Deep valleys and certain elevations are more susceptible, potentially due to freeze–thaw cycle stress [69]. Climate plays a significant role, with specific rainfall patterns linked to higher landslide incidences, likely from soil saturation [70]. Moreover, indices like the Topographic Wetness Index (TWI) and Terrain Ruggedness Index (TRI) provide insights, showing that higher moisture levels and irregular terrain surfaces promote landslides [71]. Additionally, terrain relief and steep slopes are crucial, with certain ranges more prone to landslides due to climatic impacts and gravitational forces. Proximity to active faults and roads also affects landslide occurrence, arising from geological instability and human-induced topographic changes. Terrain aspect, particularly south and west-facing slopes, and certain terrain curvatures, influence landslide susceptibility, altering sunlight exposure, moisture conditions, and water flow patterns, which can increase erosion risks.

In the field of landslide susceptibility modeling, the use of various factors can lead to complications such as overly generalized models due to high dimensionality and overfitting, which can make the results unreliable [72,73]. Each study area's distinct characteristics require a customized approach of selecting conditioning factors, aiming to minimize dimensionality and pinpoint the most important factors for that area. The need for up-to-date information is also paramount, as data from official sources often lack recent updates, affecting the precision of the models. It is important to select factors in both human-related and climatic elements to create more efficient models for managing landslide risks. Moreover, additional research is needed to analyze the importance of these factors and to refine feature selection methods, enhancing the accuracy and reducing the time and computational effort involved in LSM development and data gathering.

The study in the Himalayas' China–Pakistan Economic Corridor region highlights the importance of landslide susceptibility assessment due to unique geological and infrastructural features [17,74]. It reveals that Support Vector Machines and Logistic Regression are particularly effective in identifying high-risk areas, like those near fault lines and roads. These models excel in defining boundaries and estimating probabilities, with the Support Vector Machine also correlating well with tree-based methods and Convolutional Neural Networks, indicating their versatility in assessing various risk factors like terrain ruggedness and rainfall patterns. However, the varied correlations of Random Forest Classifier with other models point to the need for diverse approaches, especially for localized factors. This research shows the significance of strategic model selection, utilizing the strengths of different models to address the complex challenges of landslide risk assessment in this geologically sensitive region.

While our research focuses on a specific region within the Himalayas, potentially limiting the direct applicability of our findings to areas with different geological and environmental conditions, it provides valuable comparative insights that extend beyond the immediate study area. The relationships identified between factors such as slope, terrain relief, and topographic wetness index can inform similar studies in other mountainous or geologically active regions [75,76]. This comprehensive framework for integrating multiple environmental variables to predict landslide susceptibility can be adapted to different geographical settings, enhancing the accuracy of susceptibility maps when adjusted for local data. The robust performance metrics of the Random Forest Classifier and Convolutional Neural Network models in our study suggest that these models can achieve similar accuracy in other regions with appropriate adjustments. Additionally, the insights gained from our analysis, such as identifying high-risk zones and understanding the influence of various environmental factors, provide a framework for risk management and policy development that can be applied elsewhere [77,78]. Although different geological contexts may require

recalibration of the models and the quality of available data may vary by region, our work significantly contributes to the field of landslide susceptibility mapping and offers valuable guidelines for future research and policy development in diverse geographical contexts.

Landslide Susceptibility Mapping (LSM) is a critical tool for identifying areas at risk and predicting possible landslide incidents, which are significant threats to environmental stability and socioeconomic advancement. By utilizing LSM, stakeholders such as policymakers, governmental leaders, land use planners, real estate professionals, environmental groups, transportation and utility entities, agricultural and forestry sectors, insurance firms, and the general public can develop informed strategies to mitigate the adverse impacts of landslides [79,80]. This research highlights the effectiveness of LSM in pinpointing areas prone to landslides within the China–Pakistan Economic Corridor in the Himalayas, particularly high-risk zones primarily located in the central regions of our study area, along the Karakoram Highway, and in towns such as Gilgit, Jaglot, and Shandur. Recognizing these high-risk areas aids decision-makers in strategically avoiding them for agricultural, residential, and infrastructural development. To protect these susceptible areas, a combination of structural, geotechnical, political, legal, and administrative measures can be implemented to safeguard both the population and the environment from the threats posed by landslides.

5. Conclusions

This study conducted a comprehensive analysis of landslide susceptibility within the Himalayan region of the China–Pakistan Economic Corridor, utilizing high-resolution Sentinel-2 satellite imagery and Google Earth data to identify 2197 landslide polygons. The landslides were validated through literature reviews, NASA's geo-hazard catalog, and extensive field surveys, revealing that the majority of landslides, 71.13%, are very small, each covering area less than 1×10^5 m^2, while very large landslides, each covering more than 1×10^6 m^2, are the least common, representing only 0.21% of the total landslides. Non-landslide areas were delineated through ESRI ArcGIS using a random selection approach. The dataset was then divided into 70% for model training (1538 locations) and 30% for validation (659 locations), ensuring a robust model development process. Additionally, the study employed 13 conditioning factors, including slope, aspect, curvature, land use, land cover, and lithology, etc., obtained from various sources. These factors were analyzed using Pearson Correlation Coefficients, which reveal a strong positive correlation (0.92) between slope and terrain relief. This indicates that steeper slopes are prevalent in regions with significant relief, thereby impacting landslide susceptibility. Conversely, a negative correlation (-0.55) was observed between slope and the Topographic Wetness Index, suggesting lower wetness indices in steeper areas, which can influence water runoff patterns and erosion susceptibility.

Furthermore, the multicollinearity assessment revealed minimal interaction among most variables; however, elevation exhibited a higher VIF of 2.98, implying more pronounced interactions with other conditioning factors. These findings show the complexity of interactions in geographical modeling and emphasize the importance of careful selection of predictive variables in landslide susceptibility models. Moreover, machine learning models varied in their categorization of risk zones, with the Random Forest Classifier and Convolutional Neural Network categorizing larger regions as VERY LOW risk and the Support Vector Machine covering more area as LOW risk than Random Forest Classifier and K-Nearest Neighbors, highlighting the significance of model selection based on the specific characteristics and data sensitivities of the study area. These findings align with the study's aim to precisely model landslide susceptibility and highlight the need for improved prediction through dynamic variables and machine learning techniques. This research is crucial for identifying at-risk areas and predicting landslides, which threaten environmental stability and socioeconomic progress. Using landslide susceptibility mapping, stakeholders can develop strategies to mitigate impacts. Implementing structural, geotechnical, political,

legal, and administrative measures can protect populations and the environment from landslide threats.

Author Contributions: Conceptualization, H.Q.; methodology, M.U. and B.T.; software, M.U. and Y.W.; validation, M.U. and Y.W.; formal analysis, W.H., D.Y. and M.U.; investigation, M.U., Y.W., W.H. and D.Y.; resources, H.Q.; data curation, M.U. and B.T.; writing—original draft preparation, M.U.; writing—review and editing, B.T.; visualization, Y.W., W.H. and B.T.; supervision, H.Q.; project administration, H.Q.; funding acquisition, B.T. and H.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 42207384), the Key Research and Development Program of Shaanxi (Grant No. 2024SF-YBXM-669), the National Natural Science Foundation of China (Grant No. 42271078), and the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (Grant No. 2019QZKK0902).

Data Availability Statement: The Sentinel-1A data used in this study were provided by the European Space Agency (ESA), https://search.asf.alaska.edu/#/ (accessed on 2 February 2024), and the SRTM DEM was freely downloaded from the website https://earthexplorer.usgs.gov/ (accessed on 2 February 2024).

Acknowledgments: The authors thank the ESA for providing free Sentinel-1A datasets. The authors also thank the Google Earth Platform for providing the optical remote sensing images.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xie, C.; Huang, Y.; Li, L.; Li, T.; Xu, C. Detailed Inventory and Spatial Distribution Analysis of Rainfall-Induced Landslides in Jiexi County, Guangdong Province, China in August 2018. *Sustainability* **2023**, *15*, 13930. [CrossRef]
2. Chen, W.; Chen, Y.; Tsangaratos, P.; Ilia, I.; Wang, X. Combining Evolutionary Algorithms and Machine Learning Models in Landslide Susceptibility Assessments. *Remote Sens.* **2020**, *12*, 3854. [CrossRef]
3. Ullah, I.; Aslam, B.; Shah, S.H.I.A.; Tariq, A.; Qin, S.; Majeed, M.; Havenith, H.-B. An Integrated Approach of Machine Learning, Remote Sensing, and GIS Data for the Landslide Susceptibility Mapping. *Land* **2022**, *11*, 1265. [CrossRef]
4. Zhuo, L.; Huang, Y.; Zheng, J.; Cao, J.; Guo, D. Landslide Susceptibility Mapping in Guangdong Province, China, Using Random Forest Model and Considering Sample Type and Balance. *Sustainability* **2023**, *15*, 9024. [CrossRef]
5. Islam, F.; Riaz, S.; Ghaffar, B.; Tariq, A.; Shah, S.U.; Nawaz, M.; Hussain, M.L.; Amin, N.U.; Li, Q.; Lu, L.; et al. Landslide Susceptibility Mapping (LSM) of Swat District, Hindu Kush Himalayan Region of Pakistan, Using GIS-Based Bivariate Modeling. *Front. Environ. Sci.* **2022**, *10*, 1027423. [CrossRef]
6. Ali, S.; Biermanns, P.; Haider, R.; Reicherter, K. Landslide Susceptibility Mapping by Using GIS along the China–Pakistan Economic Corridor (Karakoram Highway), Pakistan. *Nat. Hazards Earth Syst. Sci.* **2018**, *19*, 999–1022. [CrossRef]
7. Habumugisha, J.M.; Chen, N.; Rahman, M.; Islam, M.M.; Ahmad, H.; Elbeltagi, A.; Sharma, G.; Liza, S.N.; Dewan, A. Landslide Susceptibility Mapping with Deep Learning Algorithms. *Sustainability* **2022**, *14*, 1734. [CrossRef]
8. Qin, Z.; Zhou, X.; Li, M.; Tong, Y.; Luo, H. Landslide Susceptibility Mapping Based on Resampling Method and FR-CNN: A Case Study of Changdu. *Land* **2023**, *12*, 1213. [CrossRef]
9. Hussain, M.A.; Chen, Z.; Zheng, Y.; Zhou, Y.; Daud, H. Deep Learning and Machine Learning Models for Landslide Susceptibility Mapping with Remote Sensing Data. *Remote Sens.* **2022**, *15*, 4703. [CrossRef]
10. Lucchese, L.V.; De Oliveira, G.G.; Pedrollo, O.C. Investigation of the Influence of Nonoccurrence Sampling on Landslide Susceptibility Assessment Using Artificial Neural Networks. *CATENA* **2021**, *198*, 105067. [CrossRef]
11. Sellers, C. Analysis of Conditioning Factors in Cuenca, Ecuador, for Landslide Susceptibility Maps Generation Employing Machine Learning Methods. *Land* **2023**, *12*, 1135. [CrossRef]
12. Band, L.E.; Moore, I.D. Scale: Landscape Attributes and Geographical Information Systems. *Hydrol. Process.* **1995**, *9*, 401–422. [CrossRef]
13. Sahin, E.K.; Colkesen, I.; Acmali, S.S.; Akgun, A.; Aydinoglu, A.C. Developing Comprehensive Geocomputation Tools for Landslide Susceptibility Mapping: LSM Tool Pack. *Comput. Geosci.* **2020**, *144*, 104592. [CrossRef]
14. Micheletti, N.; Foresti, L.; Robert, S.; Leuenberger, M.; Pedrazzini, A.; Jaboyedoff, M.; Kanevski, M. Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping. *Math. Geosci.* **2014**, *46*, 33–57. [CrossRef]
15. Lee, S.; Hong, S.M.; Jung, H.S. A Support Vector Machine for Landslide Susceptibility Mapping in Gangwon Province, Korea. *Sustainability* **2017**, *9*, 48. [CrossRef]
16. Meena, S.R.; Puliero, S.; Bhuyan, K.; Floris, M.; Catani, F. Assessing the Importance of Conditioning Factor Selection in Landslide Susceptibility for the Province of Belluno. *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 1395–1417. [CrossRef]
17. Shahzad, N.; Ding, X.; Abbas, S. A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan. *Appl. Sci.* **2022**, *12*, 2280. [CrossRef]

18. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide Susceptibility Mapping Using Random Forest, Boosted Regression Tree, Classification and Regression Tree, and General Linear Models. *Landslides* **2016**, *13*, 839–856. [CrossRef]
19. Zhang, H.; Yin, C.; Wang, S.; Guo, B. Landslide Susceptibility Mapping Based on Landslide Classification and Improved Convolutional Neural Networks. *Nat. Hazards* **2023**, *116*, 1931–1971. [CrossRef]
20. Rehman, A.; Song, J.; Haq, F.; Mahmood, S.; Ahamad, M.I.; Basharat, M.; Sajid, M.; Mehmood, M.S. Multi-Hazard Susceptibility Assessment Using the Analytical Hierarchy Process and Frequency Ratio Techniques in the Northwest Himalayas, Pakistan. *Remote Sens.* **2022**, *14*, 554. [CrossRef]
21. Ali, Z.; Qaisar, M.; Mahmood, T.; Shah, M.A.; Iqbal, T.; Serva, L.; Michetti, A.M.; Burton, P.W. The Muzaffarabad, Pakistan, Earthquake of 8 October 2005: Surface Faulting, Environmental Effects and Macroseismic Intensity. *Geol. Soc. Lond. Spec. Publ* **2009**, *316*, 155–172. [CrossRef]
22. Abid, M.; Isleem, H.F.; Shahzada, K.; Khan, A.U.; Kamal Shah, M.; Saeed, S.; Aslam, F. Seismic Hazard Assessment of Shigo Kas Hydro-Power Project (Khyber Pakhtunkhwa, Pakistan). *Buildings* **2021**, *11*, 349. [CrossRef]
23. Dhuime, B.; Bosch, D.; Bodinier, J.; Garrido, C.; Bruguier, O.; Hussain, S.; Dawood, H. Multistage Evolution of the Jijal Ultramafic–Mafic Complex (Kohistan, N Pakistan): Implications for Building the Roots of Island Arcs. *Earth Planet. Sci. Lett.* **2007**, *261*, 179–200. [CrossRef]
24. Khan, M.; Nawaz, S.; Radwan, A.E. New Insights into Tectonic Evolution and Deformation Mechanism of Continental Foreland Fold-Thrust Belt. *J. Asian Earth Sci.* **2023**, *245*, 105556. [CrossRef]
25. Rehman, S.A.; Riaz, Z.; Bui, M.D.; Rutschmann, P. Application of a 1D Numerical Model for Sediment Management in Dasu Hydropower Project. In Proceedings of the 14th International Conference on Environmental Science and Technology, Rhodes, Greece, 3–5 September 2015.
26. Hastenrath, S.; Greischar, L. The Monsoonal Current Regimes of the Tropical Indian Ocean: Observed Surface Flow Fields and Their Geostrophic and Wind-Driven Components. *J. Geophys. Res. Ocean.* **1991**, *96*, 12619–12633. [CrossRef]
27. Liu, X.; Zhao, C.; Zhang, Q.; Lu, Z.; Li, Z.; Yang, C.; Liu, C. Integration of Sentinel-1 and ALOS/PALSAR-2 SAR Datasets for Mapping Active Landslides along the Jinsha River Corridor, China. *Eng. Geol.* **2021**, *284*, 106033. [CrossRef]
28. Khan, H.; Shafique, M.; Khan, M.A.; Bacha, M.A.; Shah, S.U.; Calligaris, C. Landslide Susceptibility Assessment Using Frequency Ratio, a Case Study of Northern Pakistan. *Egypt. J. Remote Sens. Space Sci.* **2019**, *22*, 11–24. [CrossRef]
29. Fayaz, A.; Latif, M.; Khan, K. *Landslide Evaluation and Stabilization between Gilgit ans Thakot along the Karakoram Highway*; Geological Survey of Pakistan: Islamabad, Pakistan, 1985.
30. Khan, K.; Fayaz, A.; Latif, M.; Wazir, A. *Rock and Debris Slides between Khunjrab Pass and Gilgit along the Karakoram Highway*; Geological Survey of Pakistan: Islamabad, Pakistan, 1986.
31. Khan, K.; Fayaz, A.; Hussain, M.; Latif, M. *Landslides Problems and Their Mitigation along the Karakoram Highway*; Geological Survey of Pakistan: Islamabad, Pakistan, 2003.
32. Hewitt, K. Catastrophic landslides and their effects on the Upper Indus streams, Karakoram Himalaya, northern Pakistan. *Geomorphology* **1998**, *26*, 47–80. [CrossRef]
33. Hussain, M.A.; Chen, Z.; Zheng, Y.; Shoaib, M.; Shah, S.U.; Ali, N.; Afzal, Z. Landslide Susceptibility Mapping Using Machine Learning Algorithm Validated by Persistent Scatterer In-SAR Technique. *Sensors* **2022**, *22*, 3119. [CrossRef]
34. Yao, Z.; Chen, M.; Zhan, J.; Zhuang, J.; Sun, Y.; Yu, Q.; Yu, Z. Refined Landslide Susceptibility Mapping by Integrating the SHAP-CatBoost Model and InSAR Observations: A Case Study of Lishui, Southern China. *Appl. Sci.* **2023**, *13*, 12817. [CrossRef]
35. Ali, N.; Chen, J.; Fu, X.; Ali, R.; Hussain, M.A.; Daud, H.; Hussain, J.; Altalbe, A. Integrating Machine Learning Ensembles for Landslide Susceptibility Mapping in Northern Pakistan. *Remote Sens.* **2024**, *16*, 988. [CrossRef]
36. Sun, X.; Chen, J.; Li, Y.; Rene, N.N. Landslide Susceptibility Mapping along a Rapidly Uplifting River Valley of the Upper Jinsha River, Southeastern Tibetan Plateau, China. *Remote Sens.* **2022**, *14*, 1730. [CrossRef]
37. Yu, H.; Pei, W.; Zhang, J.; Chen, G. Landslide Susceptibility Mapping and Driving Mechanisms in a Vulnerable Region Based on Multiple Machine Learning Models. *Remote Sens.* **2023**, *15*, 1886. [CrossRef]
38. Minh, N.Q.; Huong, N.T.T.; Khanh, P.Q.; Hien, L.P.; Bui, D.T. Impacts of Resampling and Downscaling Digital Elevation Model and Its Morphometric Factors: A Comparison of Hopfield Neural Network, Bilinear, Bicubic, and Kriging Interpolations. *Remote Sens.* **2024**, *16*, 819. [CrossRef]
39. Stojilković, B. Towards Transferable Use of Terrain Ruggedness Component in the Geodiversity Index. *Resources* **2022**, *11*, 22. [CrossRef]
40. Winzeler, H.E.; Owens, P.R.; Read, Q.D.; Libohova, Z.; Ashworth, A.; Sauer, T. Topographic Wetness Index as a Proxy for Soil Moisture in a Hillslope Catena: Flow Algorithms and Map Generalization. *Land* **2022**, *11*, 2018. [CrossRef]
41. Cheng, J.; Dai, X.; Wang, Z.; Li, J.; Qu, G.; Li, W.; She, J.; Wang, Y. Landslide Susceptibility Assessment Model Construction Using Typical Machine Learning for the Three Gorges Reservoir Area in China. *Remote Sens.* **2022**, *14*, 2257. [CrossRef]
42. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42. [CrossRef]
43. Selamat, S.N.; Majid, N.A.; Taha, M.R.; Osman, A. Landslide Susceptibility Model Using Artificial Neural Network (ANN) Approach in Langat River Basin, Selangor, Malaysia. *Land* **2022**, *11*, 833. [CrossRef]
44. Wang, G.; Chen, X.; Chen, W. Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 144. [CrossRef]

45. Yin, Y.; Jang-Jaccard, J.; Xu, W.; Singh, A.; Zhu, J.; Sabrina, F.; Kwak, J. IGRF-RFE: A Hybrid Feature Selection Method for MLP-Based Network Intrusion Detection on UNSW-NB15 Dataset. *J. Big Data* **2023**, *10*, 15. [CrossRef]
46. Kavzoglu, T.; Teke, A.; Yilmaz, E.O. Shared Blocks-Based Ensemble Deep Learning for Shallow Landslide Susceptibility Mapping. *Remote Sens.* **2021**, *13*, 4776. [CrossRef]
47. Kamran, K.V.; Feizizadeh, B.; Khorrami, B.; Ebadi, Y. A Comparative Approach of Support Vector Machine Kernel Functions for GIS-Based Landslide Susceptibility Mapping. *Appl. Geomat.* **2021**, *13*, 837–851. [CrossRef]
48. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. A Comparative Study of Landslide Susceptibility Maps Produced Using Support Vector Machine with Different Kernel Functions and Entropy Data Mining Models in China. *Bull. Eng. Geol. Environ.* **2018**, *77*, 647–664. [CrossRef]
49. Chen, W.; Sun, Z.; Han, J. Landslide Susceptibility Modeling Using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. *Appl. Sci.* **2019**, *9*, 171. [CrossRef]
50. Taalab, K.; Cheng, T.; Zhang, Y. Mapping Landslide Susceptibility and Types Using Random Forest. *Big Earth Data* **2018**, *2*, 159–178. [CrossRef]
51. Parmar, A.; Katariya, R.; Patel, V. A Review on Random Forest: An Ensemble Classifier. In Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, Coimbatore, India, 7–8 August 2018; Springer International Publishing: Cham, Switzerland, 2019; pp. 758–763.
52. Nikoobakht, S.; Azarafza, M.; Akgün, H.; Derakhshani, R. Landslide Susceptibility Assessment by Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 5992. [CrossRef]
53. Rong, G.; Alu, S.; Li, K.; Su, Y.; Zhang, J.; Zhang, Y.; Li, T. Rainfall Induced Landslide Susceptibility Mapping Based on Bayesian Optimized Random Forest and Gradient Boosting Decision Tree Models—A Case Study of Shuicheng County, China. *Water* **2020**, *12*, 3066. [CrossRef]
54. Song, Y.; Niu, R.; Xu, S.; Ye, R.; Peng, L.; Guo, T.; Chen, T. Landslide Susceptibility Mapping Based on Weighted Gradient Boosting Decision Tree in Wanzhou Section of the Three Gorges Reservoir Area (China). *ISPRS Int. J. Geo-Inf.* **2018**, *8*, 4. [CrossRef]
55. Tang, H.; Wang, C.; An, S.; Wang, Q.; Jiang, C. A Novel Heterogeneous Ensemble Framework Based on Machine Learning Models for Shallow Landslide Susceptibility Mapping. *Remote Sens.* **2023**, *15*, 4159. [CrossRef]
56. Nayak, S.; Bhat, M.; Subba Reddy, N.V.; Ashwath Rao, B. Study of Distance Metrics on K-Nearest Neighbor Algorithm for Star Categorization. *J. Phys. Conf. Ser.* **2022**, *2161*, 012004. [CrossRef]
57. Jati, M.F.K.; Wahyunggoro, O.; Bejo, A.; Adimedha, T.B. Distance Metrics Comparison on K-Nearest Neighbor for Landslide Susceptibility Mapping. In Proceedings of the 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 13–15 July 2023.
58. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance Evaluation of the GIS-Based Data Mining Techniques of Best-First Decision Tree, Random Forest, and Naïve Bayes Tree for Landslide Susceptibility Modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [CrossRef]
59. Raia, S.; Alvioli, M.; Rossi, M.; Baum, R.L.; Godt, J.W.; Guzzetti, F. Improving Predictive Power of Physically Based Rainfall-Induced Shallow Landslide Models: A Probabilistic Approach. *Geosci. Model. Dev.* **2014**, *7*, 495–514. [CrossRef]
60. Vakhshoori, V.; Zare, M. Is the ROC Curve a Reliable Tool to Compare the Validity of Landslide Susceptibility Maps? *Geomat. Nat. Hazards Risk* **2018**, *9*, 249–266. [CrossRef]
61. Lei, T.; Zhang, Y.; Lv, Z.; Li, S.; Liu, S.; Nandi, A.K. Landslide Inventory Mapping from Bitemporal Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 982–986. [CrossRef]
62. Adugna, T.; Xu, W.; Fan, J. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sens.* **2022**, *14*, 574. [CrossRef]
63. Datta, A.; Luirei, K.; Mehta, M. Slope Instabilities and Evolution of Tectonic Geomorphology Along the Strike of the Main Boundary Thrust Zone in the Western Himalaya, India. *Nat. Hazards Res.* **2023**, *4*, 118–133. [CrossRef]
64. Qiu, H.; Su, L.; Tang, B.; Yang, D.; Ullah, M.; Zhu, Y.; Kamp, U. The Effect of Location and Geometric Properties of Landslides Caused by Rainstorms and Earthquakes. *Earth Surf. Process. Landf.* **2024**, *49*, 2067–2079. [CrossRef]
65. Qiu, H.; Zhu, Y.; Zhou, W.; Sun, H.; He, J.; Liu, Z. Influence of DEM Resolution on Landslide Simulation Performance Based on the Scoops3D Model. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1663–1681. [CrossRef]
66. Shao, X.; Xu, C.; Ma, S. Preliminary Analysis of Coseismic Landslides Induced by the 1 June 2022 Ms 6.1 Lushan Earthquake, China. *Sustainability* **2022**, *14*, 16554. [CrossRef]
67. Ye, B.; Qiu, H.; Tang, B.; Liu, Y.; Liu, Z.; Jiang, X.; Yang, D.; Ullah, M.; Zhu, Y.; Kamp, U. Creep Deformation Monitoring of Landslides in a Reservoir Area. *J. Hydrol.* **2024**, *632*, 130905. [CrossRef]
68. Pandita, S.; Kumar, V.; Dutt, H.C. Environmental Variables Vis-à-Vis Distribution of Herbaceous Tracheophytes on Northern Sub-Slopes in Western Himalayan Ecotone. *Ecol. Process* **2019**, *8*, 45. [CrossRef]
69. Yang, D.; Qiu, H.; Ye, B.; Liu, Y.; Zhang, J.; Zhu, Y. Distribution and Recurrence of Warming-induced Retrogressive Thaw Slumps on the Central Qinghai-Tibet Plateau. *J. Geophys. Res. Earth Surf.* **2023**, *128*, e2022JF007047. [CrossRef]
70. Liu, Y.; Qiu, H.; Kamp, U.; Wang, N.; Wang, J.; Huang, C.; Tang, B. Higher Temperature Sensitivity of Retrogressive Thaw Slump Activity in the Arctic Compared to the Third Pole. *Sci. Total Environ.* **2024**, *914*, 170007. [CrossRef] [PubMed]

71. Saleem, N.; Huq, M.E.; Twumasi, N.Y.D.; Javed, A.; Sajjad, A. Parameters Derived from and/or Used with Digital Elevation Models (DEMs) for Landslide Susceptibility Mapping and Landslide Risk Assessment: A Review. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 545. [CrossRef]
72. Pes, B. Learning from High-Dimensional and Class-Imbalanced Datasets Using Random Forests. *Information* **2021**, *12*, 286. [CrossRef]
73. Can, R.; Kocaman, S.; Gokceoglu, C. A Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey. *Appl. Sci.* **2021**, *11*, 4993. [CrossRef]
74. Luyao, W.; Haijun, Q.; Wenqi, Z.; Yaru, Z.; Zijing, L.; Shuyue, M.; Dongdong, Y.; Bingzhe, T. The post-failure spatiotemporal deformation of certain translational landslides may follow the pre-failure pattern. *Remote Sens.* **2022**, *14*, 2333. [CrossRef]
75. Dai, X.; Zhu, Y.; Sun, K.; Zou, Q.; Zhao, S.; Li, W.; Hu, L.; Wang, S. Examining the Spatially Varying Relationships between Landslide Susceptibility and Conditioning Factors Using a Geographical Random Forest Approach: A Case Study in Liangshan, China. *Remote Sens.* **2023**, *15*, 1513. [CrossRef]
76. Zhou, W.; Qiu, H.; Wang, L.; Pei, Y.; Tang, B.; Ma, S.; Yang, D.; Cao, M. Combining rainfall-induced shallow landslides and subsequent debris flows for hazard chain prediction. *Catena* **2022**, *213*, 106199. [CrossRef]
77. Shah, N.A.; Shafique, M.; Ishfaq, M.; Faisal, K.; Van der Meijde, M. Integrated Approach for Landslide Risk Assessment Using Geoinformation Tools and Field Data in Hindukush Mountain Ranges, Northern Pakistan. *Sustainability* **2023**, *15*, 3102. [CrossRef]
78. Liu, Z.; Qiu, H.; Zhu, Y.; Liu, Y.; Yang, D.; Ma, S.; Zhang, J.; Wang, Y.; Wang, L.; Tang, B. Efficient Identification and Monitoring of Landslides by Time-Series InSAR Combining Single- and Multi-Look Phases. *Remote Sens.* **2022**, *14*, 1026. [CrossRef]
79. Alem, G. Urban Plans and Conflicting Interests in Sustainable Cross-Boundary Land Governance, the Case of National Urban and Regional Plans in Ethiopia. *Sustainability* **2021**, *13*, 3081. [CrossRef]
80. Zhu, Y.; Qiu, H.; Yang, D.; Liu, Z.; Ma, S.; Pei, Y.; He, J.; Du, C.; Sun, H. Pre- and post-failure spatiotemporal evolution of loess landslides: A case study of the Jiangou landslide in Ledu, China. *Landslides* **2021**, *18*, 3475–3484. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.