Technical Note

Landslides
DOI 10.1007/s10346-025-02532-9
Received: 12 September 2024
Accepted: 24 April 2025

Chongcai Xu · Congchao Bian · Teng Yu · Chenchen Qiu

Estimation of landslide volume by machine learning and remote sensing techniques in Himalayan regions

Abstract Topographical and geological conditions are typically regarded as the primary causes of landslides. However, accurately estimating landslide volumes on rock slopes using empirical equations remains challenging. In contrast, data science approaches, such as machine learning, leverage advanced data integration and processing capabilities, significantly enhancing the accuracy and reliability of landslide volume estimations. As such, an resemble method, XGBoost, was chosen in our study to estimate the potential landslide volume in Gyirong, China. A factor combination was proposed in this study. They are related to geomorphic (area of slope units (*S*) and mean elevation of slope units (*El*)) and geological (faults density (*Fd*) and geological index (*GI*)) conditions. The performance of the developed model was compared with three other machine learning models, including gradient boosting (GBDT), adaptive boosting (AdaBoost) and random forest (RF) based on the mean absolute percentage error (MAPE) and determination of coefficient (*R*-squared). The results demonstrate that XGBoost achieves the highest prediction accuracy, with an *R*-squared value of 0.986, and a mean absolute percentage error (MAPE) reduces to 8.19%. Additionally, the prediction outcomes of the machine learning model, using the proposed factor combination, were compared with several empirical models. Once again, XGBoost model exhibits the lowest error relative to measured values, highlighting the superiority of machine learning in landslide volume prediction and validating the effectiveness of the selected factors. Overall, the accurate estimation of landslide volume in remote areas can benefit the disaster management and decrease losses of human lives and properties.

Keywords Landslide volume estimation · Rock slopes · Machine learning model · Satellite images · Lithology

Introduction

Landslides can cause severe damage to infrastructure and loss of human lives due to the movement of vast quantities of rock down slopes under the influence of gravity (Lacasse et al. 2009; Zhao et al. 2024). Furthermore, the accumulation of crushed rock on slopes serves as a potential source of material for future debris flows (Blahut et al. 2010; Qiu et al. 2024), leading to secondary damage to infrastructure and further degradation of the ecological environment. (Petrakov et al. 2007; Jomelli et al. 2015; Perov et al. 2017). It was reported from the Center for Research on Epidemiology of Disasters (CRED) that, all over the world, almost 17% of fatalities caused by natural disasters was caused by landslide. The mountainous areas are taking 36% of the earth land and over 10% of population that are still living in these areas (Gerrard 1990). It is still a challenge to provide an early warning for landslides to protect properties and lives and mitigate the loss (Singhroy 2009). This is because landslide warning normally requires a better understanding of the rock types and orientation of bed rocks (Guzzetti et al. 2009; Gaziev 2013). Furthermore, climate change in the past several decades (IPCC, 2013) leads to an increasing numbers of extreme weather event (O'Gorman 2015; Qiu et al. 2022), which also increased the occurrence frequency of landslide events (Kirschbaum and Adler 2013).

Accurate susceptibility maps of landslide can be helpful to provide timely warning for local people and therefore reduce casualties (Stanley and Kirschbaum 2017; Chen and Li 2020). The prediction of landslide volume is a crucial component and an essential supplement to landslide risk assessment (Lacasse et al. 2009). However, accurately determining landslide volume is challenging, as it is closely related to the internal structure and geomorphic conditions of rock slopes (Guzzetti et al. 2009). Empirical relationships between landslide areas (A_L) and landslide volumes (V_L) have been established by Imaizumi and Sidle (2007), Imaizumi et al. (2008), and Guzzetti et al. (2009) based on a large number of historical landslides across the world. Qiu et al. (2017) and Zhang et al. (2020) further optimised the proposed equations and improved prediction accuracy related to loess slides using the combination method of empirical model and remote sensing techniques. However, limitations persist in using newly improved formulas to predict rock landslide volume, as the failure mechanisms of rock slopes differ significantly from those of soil slopes (Piteau and Peckover 1978). Due to the limitations from the traditional empirical method, in recent years, few studies focus on the improvement of the prediction of landslide volume for rock slopes.

To address the limitations of empirical methods and enhance the accuracy of landslide volume prediction for rock slopes in mountainous regions, this study introduces machine learning as an innovative research approach. Through support vector machine (SVM) (Tien Bui et al. 2016), logistic regression (LR) (Budimir et al. 2015; Zhou et al. 2018), AdaBoost (Adaptive Boosting) (Kadavi et al. 2018), and random forest (RF) (Youssef et al. 2016; Arabameri et al. 2020) method, machine learning has been widely applied to landslide susceptibility analysis (Marjanović et al. 2011; Di et al., 2020) and risk assessment (Pourghasemi et al. 2018; Novellino et al. 2021). The application of machine learning models has demonstrated their superiority over empirical methods in hazard analysis (Rahmati et al. 2017). However, despite the advancements in machine learning, there has been limited progress in improving the prediction accuracy of landslide volume when considering geomorphic and geological

Published online: 08 May 2025

factors. Therefore, in this paper, geomorphic and geological features extracted from 47 landslides in Gyirong were considered through the following steps.

(1) Generated slope units of identified landslides through DEM (resolution of 12.5 m and download on ASF Data Search (https://search.asf.alaska.edu/#/)) using ArcGIS.

(2) Extracted values of generated units in relation to geomorphic, geological and ecological conditions.

(3) Further decided the controlling factors of landslide volume based on the correlation analysis between each factor and identified landslide volume using Person's coefficient (r) and p-value.

(4) An ensemble model, XGBoost, was used to predict landslide volume.

(5) The boosting-related and bagging-related algorithms, AdaBoost, GBDT and RF, were introduced to evaluate the performance of XGBoost based on MAPE (Mean Absolute Percentage Error) and R-squared values (Coefficient of Determination).

(6) The prediction results of machine learning model were compared with the outcomes of selected existing empirical equations.

Overall, the developed machine learning model in this paper can be used to predict landslide volume by considering geological and geomorphic factors.

Characteristics of engineering geology in this area

Geological and geomorphic conditions in Gyirong

Gyirong was selected as the study area (Fig. 1), which is located in the southwestern region of Tibet. This area comprises six towns and spans a total area of approximately 9300 km^2. The northern part of Gyirong falls within a semi-arid plateau climate zone characterised by seasonal winds, resulting in an annual average temperature of 2 °C. The arid climate limits precipitation to 300–600 mm per year and contributes to the widespread presence of weathered rocks with low compressive and shear strength. As illustrated in Fig. 1, a total of 82 landslides were documented in the past 5 years. However, the precise volume of some landslides could not be accurately determined due to vegetation growth, water erosion and sediment migration. Consequently, 47 landslides were selected for analysis in this study. Notably, over 70% of the identified landslides occurred in the southern part of Gyirong, primarily due to abundant rainfall and fault activity, which have played a significant role in triggering landslides. However, the most critical factor influencing slope stability is the geological structure and lithology, as these determine the integrity of the internal slope structure and the location of potential sliding surface.

In terms of the characteristics of geological structure in this area, the collision and squeezing between Indian plate and Eurasian plate (Fig. 2) give birth to the Himalayas and also form a series of

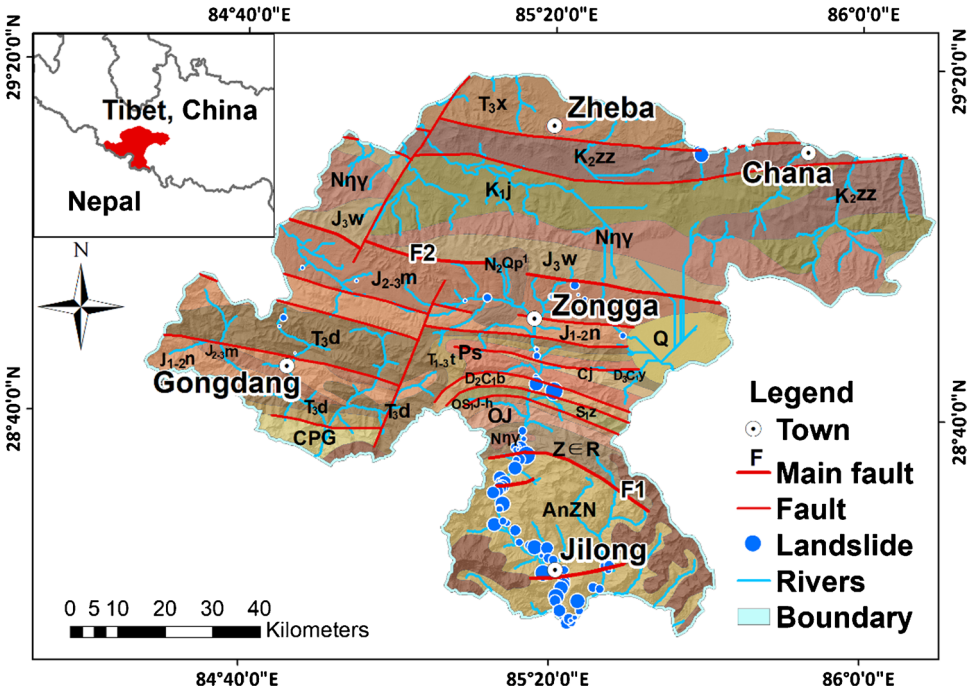


Fig. 1 Location of study area and its geological settings (D_3C_1y: limestone, dolomite, quartz sandstone, and shale; T_3x: shale, slate, metasandstone, and quartz sandstone; CPG: clastic rock, tillite, slate, and conglomerate; $T_{1-3}t$: bioclastic rock, shale, and sandy limestone; Cj: diamicite and quartz sandstone; K_2zz: calcareous shale, siliceous shale, shale, and sandstone; T_3d: quartz sandstone and fine-grained conglomerate; Nηγ: granite, plagioclase, and potash feldspar; N_2Qp^1: glutenite, sandstone, clay rock, and conglomerate; D_2C_1b: quartz sandstone, limestone, dolomite, and shale; K_1j: siliceous shale; OJ: limestone and siltstone; os_1J-h: carbonatite, calcareous siltstone, and shale; S_1z: sandstone and shale; Q: sediment, gravel, and rock fragment; J_3w: shale, quartz sandstone; AnZN: schist, gneiss, granulite, and migmatite; $J_{1-2}n$: limestone, arenaceous shale, and sandstone; Z∈R: quartz schist, slate, and phyllite; Ps: quartz sandstone, sandstone, siltstone; J_{2-3} m: shale, sandstone, silty shale, siltstone)

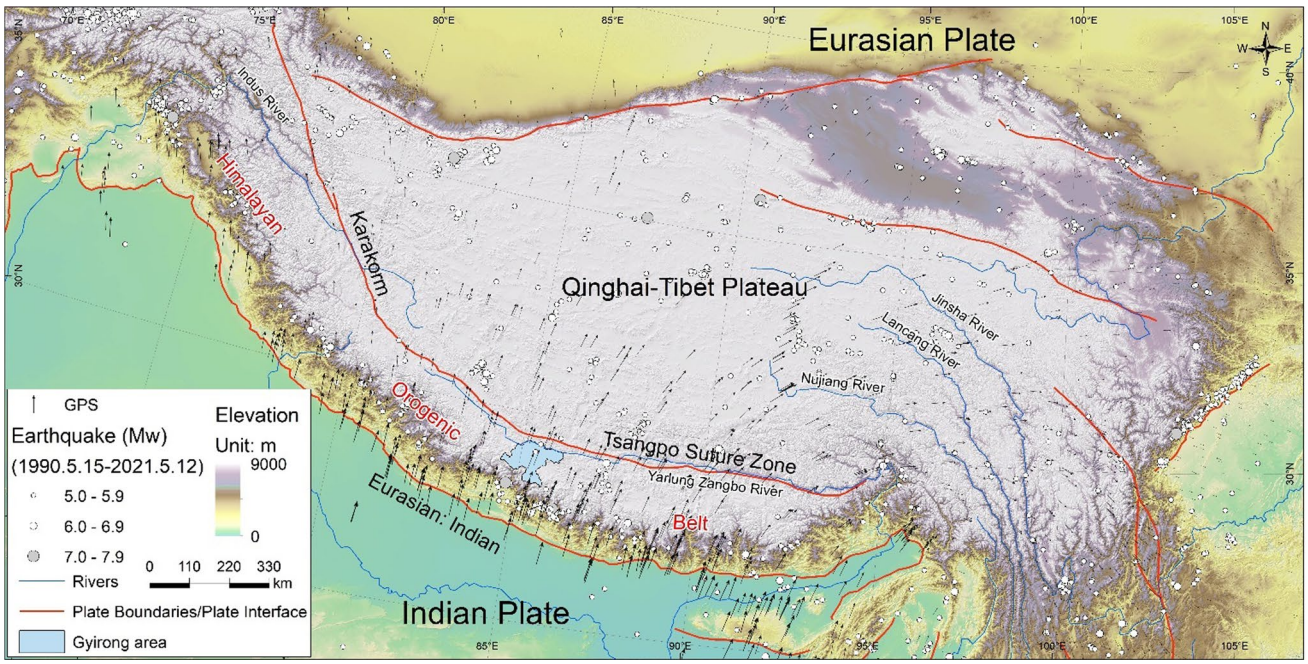


Fig. 2 Geological location of study area in the Himalayas. The GPS data is from Wang and Shen. (2020), the Plate boundaries/Plate interface and historical earthquakes are from USGS (https://www.usgs.gov/media/files/plate-boundaries-kmz-file, https://earthquake.usgs.gov/earthquakes/map)

large-scale faults (Fig. 1). These faults zones not only impact the trend and flow of rives but also control the development and distribution of natural hazards. As a result, landslide frequently occurred in this area (Fig. 3). The studies of landslide in this area can benefit the recognition and mitigation of landslide in the Himalayan regions.

In order to better clarify the geological conditions in this area, the rocks can be classified into four groups (Table 1) based on the material composition, geological structure, surface weathering degree and mechanical properties of exposed rocks and stratum.

The formation of landslides is influenced by geomorphological and geological structures, stratigraphic lithology and various external driving forces. In general, geographic and geomorphic conditions determine the stress distribution on free surfaces, making steep slopes and highly weathered slopes more susceptible to landslides. Additionally, lithology plays a crucial role in slope stability. Variations in rock lithology influence the degree of weathering resistance, leading to differences in the formation of weathered residual slope deposits. Furthermore, factors such as fold morphology, bedrock orientation and joint development significantly impact slope stability. Given these considerations, the correlations between rock structure development and landslide occurrence in this area are analysed in the "Correlation between lithology and happening of landslide" section.

Fig. 3 Identified landslide in Gyirong during field investigation

Table 1 Categories of engineering geology groups in Gyirong

Categories of engineering geology for rocks and soil				Geological code	Geomorphic unit	Characteristics of engineering geology	Uniaxial compressive strength
Soil and rocks	Rocks	Lithology	Rock types				
		Granite, porphyritic granite, siliceous rock and monzonitic granite	Solid rocks	T_3Xk γ_6^1,γ_6^2	Mountains	Compact structure and no weak intercalated layers; Slightly weathered and joints developed	> 60 Mpa
		Limestone and sandstone	Soft rocks	Oij, O_2, Ssh, C_2, D_1l, $D_{2-3}dg$, C_1, P_1, T_3x, T_3, J_3m, K	Mountains	High shear strength, moderate weathered and primary joints developed	> 30 Mpa
		Phyllite, gneiss, quartz schist, slate, shale and marlstone	Weak rocks	Anc^1, T_3g, J_3l, J_1r	Mountains	Low shear strength, poor resistance to weathering; soften by water and collapse after losing water	< 5 Mpa
	Soil	Alluvial deposit, sandy pebble soil, gravelly soil, fragmentary breccia and stone soil	Multiple-layer soil	Q	Valleyside slopes	Quaternary loose deposits, poor mechanical properties	\

Correlation between lithology and occurrence of landslide

The correlation between the distribution characteristics of lithology and landslide occurrence was further discussed in this section with the consideration of all the 82 identified landslides in Gyirong. Table 2 illustrates the litho-stratigraphic classes in this study area, in which.

Approximately 77.44% of lithology class in this area belongs to Mesozoic era, in which 31.26% is Cretaceous (K_2zz and K_1j) in the north part of this area, mainly tending direction of East–West (EW). The rocks of Cretaceous are mainly shale and sandstone interbedded with limestone. Shale has the characteristics of fragility and erodibility. Therefore, the combined influence of strong eolian activity on the plateau surface and active fault movements has led to the settlement and accumulation of extensive deposits of crushed rock on the slopes. These accumulated rock fragments serve as a potential material source for slope debris flows, which can be triggered by water flow. However, there was no landslide observed in this region. This might be due to the formation of low mountains and hills with moderate slop, and moderate slope decreases the flow kinetic energy and prevents mass movement. Another reason behind this might be the existence of shale with limestone interlayer that benefits the stability of slopes and avoids the development of sliding surface. This is because the particles of limestone are closely packed and therefore result in high compaction strength.

Furthermore, 14 landslides are identified in the period of Jurassic. The Jurassic era is mainly distributed in central part of this study area (see $J_{1-2}N$, J_{2-3} m, J_3w in Fig. 1). Specifically, 6 landslides are distributed in the bedding slopes with the lithology of shale, sandstone, silty shale and siltstone interbedded with mudstone ($J_{1-2}N$). Mudstone is classified as a weak rock. Consequently, rainfall, combined with the compressive effects of fault activity, can contribute to the formation of sliding surfaces, finally triggering slope failures. During the Late Jurassic period, 8 landslides are identified in this area, which is predominantly covered by mudstone and siltstone (J_{2-3} m). Due to the loose structure of mudstone and the widespread presence of bedding slopes, this geological formation exhibits extreme instability.

In addition to the period of Cretaceous and Jurassic, 55% of landslides are observed in the period of Middle Triassic to Late Triassic. The area is distributed in the south part with the lithology

Table 2 Areal distribution of the litho-stratigraphic classes

Era	Period	Lithology class	Area/km²	Percentage/%	Number of landslide
Cenozoic	Quaternary	Sediment, gravel, rock fragment	221.88	2.53	0
Cenozoic	Tertiary	Granite, plagioclase, potash feldspar	827.68	9.43	1
Cenozoic	Pliocene to Pleistocene	Glutenite, sandstone, clay rock, conglomerate	96.53	1.10	0
Mesozoic	Early Cretaceous to Late Cretaceous	Calcareous shale, siliceous shale, shale, sandstone interbedded with limestone	982.96	11.20	0
Mesozoic	Early Cretaceous	Siliceous shale interbedded with limestone and sandstone	1141.91	13.01	0
Mesozoic	Late Jurassic	Shale intercalated with quartz sandstone and interbedded with shale, siltstone	547.96	6.24	0
Mesozoic	Late Jurassic	Mudstone, siltstone interbedded with sandstone	898.82	10.24	8
Mesozoic	Early Jurassic to Middle Jurassic	Limestone, arenaceous shale, sandstone interbedded with mudstone	667.99	7.61	6
Mesozoic	Triassic	Clastic rock, tillite, slate, conglomerate	116.57	1.33	0
Mesozoic	Late Triassic	Shale, slate, metasandstone, quartz sandstone interbedded with marlstone, siliceous rock	486.98	5.55	1
Mesozoic	Late Triassic	Quartz sandstone interbedded with fine-grained conglomerate and limestone	750.73	8.56	5
Proterozoic	Middle Triassic to Late Triassic	Schist, gneiss, granulite, migmatite	1045.4	11.91	45
Mesozoic	Early Triassic to Late Triassic	Bioclastic rock, shale, sandy limestone interbedded with limestone, marlstone, siltstone	157.28	1.79	2
Late Paleozoic	Middle Permian to Late Permian	Quartz sandstone interbedded with shale; sandstone, siltstone interbedded with limestone	215.25	2.45	3
Late Paleozoic	Early Carboniferous to Early Permian	Diamictite interbedded with conglomerate, shale, slate; Quartz sandstone	54.14	0.62	0
Late Paleozoic	Early Carboniferous	Limestone, dolomite, shale; Quartz sandstone, shale	24.32	0.28	0
Late Paleozoic	Middle Devonian to Late Devonian, Early Carboniferous	Quartz sandstone; Limestone, dolomite, shale	108.30	1.23	2
Early Paleozoic	Early Silurian	Sandstone, shale interbedded with marlstone	96.10	1.10	0
Early Paleozoic	Early Ordovician to Late Ordovician	Limestone interbedded with calcareous siltstone	94.32	1.07	2
Early Paleozoic	Early Ordovician to Late Ordovician, Late Ordovician	Carbonatite, calcareous siltstone, arenaceous shale	54.48	0.62	0
Paleozoic	Cambrian and Sinian	Quartz schist interbedded with phyllite	185.21	2.11	7

Note: 'Area' in this table indicates the total area of each lithology class, and 'Percentage' demonstrates the ratio of area of each lithology class and the total area of this study region. The era of identified rocks ranges from Neoproterozoic to Cenozoic, which also indicates the complex geological activities and appropriate hazard-inducing environment

of schist, gneiss, granulite and migmatite (A_nZN). Unlike the failure mechanism of weak intercalated layers, schist and gneiss, though classified as metamorphic rocks, exhibit a lamellar structure with characteristics of low compressive and shear strength along the direction of the lamellar planes, making them more susceptible to deformation and failure.

Similarly, 7 landslides occurred in the period of Sinian, the lithology group is decided as quartz schist, slate, phyllite interbedded with limestone and metasandstone ($Z \in R$). The primary vertical joint within the quartz schist and phyllite promoted the development of weak structural surface and finally can induce landslide. Moreover, 5 landslides are identified in the period of Late

Triassic with the lithology of quartz sandstone interbedded with fine-grained conglomerate and limestone (T_3d). A well-developed vertical joint structure in quartz sandstone creates natural channels for water flow. As a result, the long-term effects of down-cutting and erosion weaken the structural integrity of the rock, ultimately leading to landslides. As for the other identified landslides in period of Middle Permian to Late Permian, Middle Devonian and Early Ordovician to Late Ordovician, the existence of weak intercalated layer is the main cause. Rainfall softens the weak layers and therefore decreases the sliding resistance of sliding surface, inducing landslide.

Overall, this study area is situated in the central region of the Himalayas and features a diverse landscape, including plateaus, depressions and alpine valleys extending from north to south. The compressional forces between the Indian Plate and the Eurasian Plate have resulted in complex geomorphic and geological conditions, which contribute to the occurrence of landslides. To predict potential landslide volumes, geological and geomorphic factors are analysed in the following sections, with a particular focus on the influence of lithology and rock orientation on landslide volume.

Methodology

In this paper, an ensemble machine learning algorithm, XGBoost, was applied to achieve the prediction of the potential accumulated volume of one landslide in Gyirong. After that, several methods, RF, GBDT and AdaBoost, were introduced to evaluate the performance of XGBoost in estimating the possible magnitude of landslide based on MAPE (Mean Absolute Percentage Error) and R-squared. The workflow in this paper is presented in Fig. 4.

The details of workflow comprise the following steps:

(1) Landslide identification.

The occurred landslides in the past five years were identified based on historical satellite images and field investigations, and the accumulated volume of one landslide was also decided.

(2) Generation of slope units.

There are five methods used for evaluation of slope stability (Guzzettie et al., 1999), including grid cell, terrain cell, unique-condition unit, slope unit and topographic unit. The selection of evaluation unit is the foundation of regional evaluation of slope stability, which also decides the accuracy of prediction result of landslide magnitude. Indeed, grid unit (Apip et al. 2010; Muntohar and Liao 2009; Liao et al. 2011) and slope unit (Giles and Franklin 1998; Xie et al. 2004) are two most widely used methods for landslide analysis, while grid cell units cannot provide accurate location of the potential landslide and reflect topographic features of slope failure when compared with slope units. Additionally, slope units can reflect the complete surface geometry of slopes which are prone to landslide and avoid difficulty in deciding landslide area due to vegetation coverage (Sun et al., 2020). Therefore, slope unit method is usually used for the prediction and evaluation of landslide (Alvioli et al. 2014).

As shown in Fig. 5, the slope units in this paper were extracted based on the 'Hydrology' tool in GIS. After that, the extracted units were imported into Google Earth for further revision to ensure the homogeneity of slop gradient and aspect within one slope unit. The part of final results of slope unit extraction were presented in Fig. 5.

(3) Controlling factors selection and correlation analysis.

A total of seven factors in relation to geomorphic, geological and ecological conditions were selected to predict the volume of potential landslide based on the identified 47 historical landslides in the past 5 years, including area of slope unit, mean curvature of slope unit, mean elevation of slope units, geological index, mean

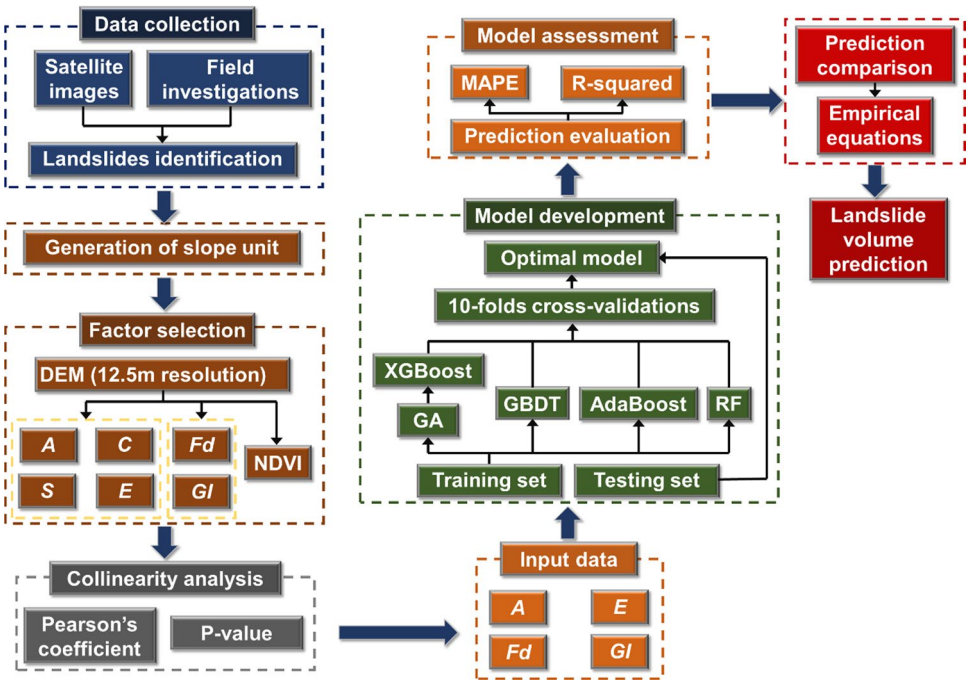


Fig. 4 Schematic flowchart of this study

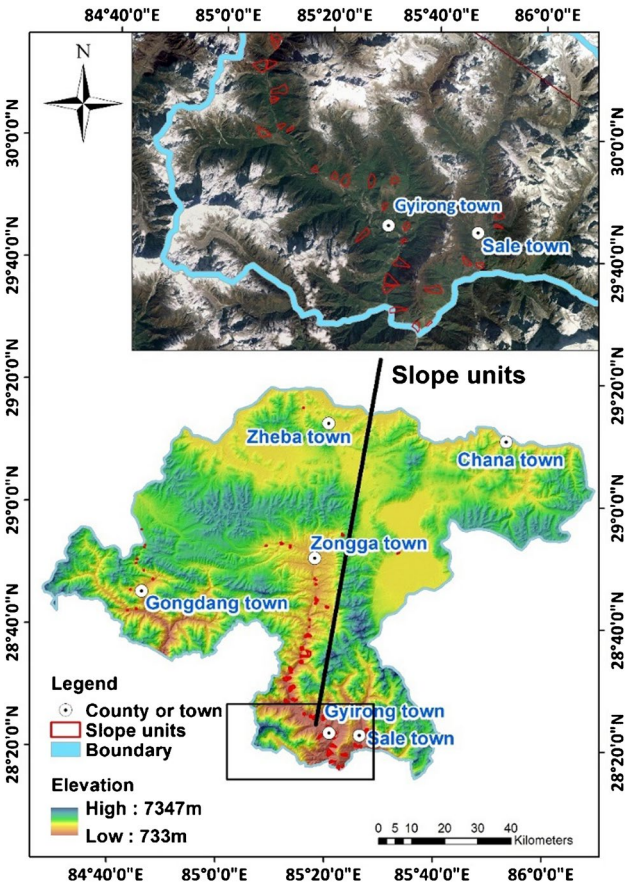


Fig. 5 Division of slope units

fault intensity of slope unit and mean NDVI (MODIS data (https://modis.gsfc.nasa.gov/)) (see Table 3).

a)Geomorphic and ecological factors.

There were four factors selected in geomorphic group to indicate the potential volume of landslide, including area of slope unit (S), mean slope gradient of slope unit (Sl), mean curvature of slope unit (Slc) and mean elevation of slope units (El).

The correlation between landslide area and volume had been indicated in Simonett (1967) and Hovious et al. (1997) even though landslide area in these papers represented total area of the failure and sediment. While in order to reflect the whole processes of initiation, propagation and accumulation of landslide (Martinello et al. 2021), the whole area of slope units was preferred in this paper to estimate the potential volume of landslide. Another topographic factor, mean slope gradient of slope unit (Sl), was used to reflect the surface fluctuation and roughness of slopes (Tanyaş et al. 2017) as well as mean curvature of slop units (Slc) (Tanyaş et al. 2019) since steep slope could cause serious gravity erosion and increase soil erosion intensity due to the high flow kinetic energy, which might decide the weak structural surface of slopes. Moreover, mean elevation of slope units (El) can reflect the degree of weathering for rocks since the higher altitude may result in greater weathering effects and generate more loose materials.

Normalised density vegetation index (NDVI), belonging to the ecological group, was used to represent the vegetation coverage in this area. The slope failure can be interpreted as the decreasing of effective stress within soil and the increasing of excess pore water pressure, while the roots could to some extent mitigate the failure process through anchoring and reinforcement effects. Vegetation roots can work as effective anchor to enhance the soil cohesion, and therefore increase shear strength of the composite soil and the safety factor can be increased by at least 8% (Ali et al., 2012). Therefore, the depth of vegetation roots can impact the locations of potential sliding surface and further result in the variation of landslide magnitude.

b) Geological factors.

Geological index (GI) in Table 4 was used to quantify the erodibility of the exposed rocks based on (D' Agostino and Marchi 2001). A slope with higher erodibility rocks could result in a larger volume of potential landslide when compared with slope with lower erodibility rocks.

For basins entirely or almost entirely consisting of this lithology, a cautionary value of 0.5 is advised.

As for the impacts of faults movement on internal structure of rock slopes, fault density (F_d) in this paper was firstly proposed to reflect the structure stability of slopes and help to estimate potential landslide volume. As indicated in Fig. 1, there are two main faults crossing this area except the other several secondary faults, named Todan-Neira fault (F1) and Chimazun-Duoqingcuo fault (F2). The F1 separates the North Himalayan structural belt and the Higher Himalayan tectonic belt, presenting lithology of Sinian and Cambrian in the north part and Presinian in the south part. The

Table 3 Selected factors which may impact the magnitude of landslide

Categories	Factors	Unit
Geomorphic factors	Area of slope unit (S)	km^2
	Mean slope gradient of slope unit (Sl)	°
	Mean curvature of slope unit (Slc)	\
	Mean elevation of slope units (El)	m
Geological factors	Geological index (GI)	\
	Fault density (Fd)	\
Ecological factors	Normalised Density Vegetation Index (NDVI)	\

Table 4 Lithology classes and geological index

Lithology class	GI value
Quaternary deposits	5
Schists and phyllites	4
Marls, marly limestone, siltstones, etc	3
Volcaniclastic rocks	2
Dolomite and limestone rocks	1
Massive igneous and metamorphic rocks	0
Intensely fractured and weathered rocks for all lithologies	3–5

main section of F1 leans to the north at a changing angle, ranging from 10 to 65°. As a result, the active fault movement caused crushed rocks and fold belts due to intense and long-term compression. While the invasion of granite, gneiss and white mica were observed in the south side because of the same geological activity. Additionally, the F2 fault stretches from north-west (NW) direction to near east–west (EW) direction gradually. Similarly, the main section leans to the north at an angle between 30 and 50°, controlling lithology variation of marine sediment.

(4) Data normalisation and collinearity analysis.

In order to avoid overfitting when developing machine learning model, stability of input data is essential to decrease the possibility of local optimisation and minimise the impacts of noise data on final results. Therefore, collinearity analysis is essential for the selected seven factors before model training. The independent and dependent variables in Table 5 were rescaled into the range between 0.01 and 0.99 based on the following equation (Lee et al. 2021):

$$x_{nor} = \frac{x - \min(x)}{\max(x) - \min(x)}(U - L) + L \quad (1)$$

where x_{nor} represents the normalised value and x is the original value. U and L represent the upper and lower normalisation bounds, respectively. After the completion of data normalisation, correlation analysis between each controlling factor and volume was conducted to investigate the impact of these indicators on landslide volume based on Pearson's coefficient (r) and p-value using SPSS statistics software.

Pearson's coefficient was used to describe the linear relationship between input data and landslide volume so as to exclude the factors that may exhibit a weak correlation with landslide volume. This is because the effect of introducing noise may outweigh their contributions to model development when these factors are incorporated into model training. As for the employment of p-value ($p < 0.05$) in this study, this index excels in investigating the significance between dependent and independent variables. In detail, a low p-value ($p > 0.05$) indicates that the correlation is statistically significant. However, a high p-value implies insufficient evidence to reject the null hypothesis of no correlation, which suggests that the observed correlation might be random correlations. Therefore, the coupled usage of r and p-value can ensure that the observed

Table 5 Collected data of the controlling factors

No	S/km²	Fd	GI	El/m	$V/10^4 m^3$
1	0.3411	0.1276	3.5	4770	3.9
2	0.3096	0.1140	3.5	4712	3.1
3	0.0832	0.0866	3.5	3998	0.8
4	0.4031	0.1535	3.5	4212	5.6
5	0.1483	0.0328	3.5	5028	1.5
6	0.2125	0.0831	4.5	4624	2
7	0.3681	0.1808	4.5	4768	5.3
8	0.0927	0.0988	3.5	5162	0.8
9	0.3844	0.1664	3.5	5014	5
10	0.0842	0.0968	3.5	4740	0.5
11	0.2383	0.1125	0.5	4028	2.1
12	0.7459	0.1299	4.5	4130	7.5
13	0.0738	0.0923	0.5	3629	0.5
14	0.1242	0.0887	0.5	3653	0.66
15	0.4956	0.0834	0.5	3718	3.6
16	0.2328	0.0611	0.5	3578	2.24
17	0.1407	0.0404	0.5	2993	1.344
18	0.0612	0.0382	0	2876	0.225
19	0.2189	0.0409	0.5	3100	2.052
20	0.0410	0.0387	0.5	3341	0.1
21	0.0850	0.0307	0.5	2539	0.5
22	0.0558	0.0304	0.5	2619	0.25
23	0.1010	0.0287	4.5	2526	0.535
24	0.1100	0.0194	0.5	2292	0.3
25	0.0453	0.0179	0.5	2219	0.1
26	0.2499	0.0196	0.5	3004	2.5
27	0.1636	0.0174	0.5	2282	1.5
28	0.0923	0.0154	0.5	2077	0.6
29	0.5954	0.0155	0.5	2175	0.75
30	0.0708	0.0154	0.5	2121	0.66
31	0.0849	0.0157	0.5	2131	0.5
32	0.1554	0.0157	0.5	2259	1.1
33	0.0813	0.0281	0.5	3057	0.5
34	0.3756	0.0295	0.5	3223	5
35	0.0412	0.0368	0.5	3319	0.15
36	0.0248	0.0369	0.5	3184	0.34
37	0.0433	0.0389	0.5	3130	0.25

Table 5 (continued)

No	S/km²	Fd	GI	El/m	V/10⁴m³
38	0.0430	0.0397	0.5	2784	0.12
39	0.0665	0.0398	0.5	2690	0.32
40	0.0514	0.0375	0.5	2056	0.19
41	0.0403	0.0372	0.5	1943	0.17
42	0.0346	0.0362	0.5	1721	0.12
43	0.0315	0.0357	0.5	1848	0.1
44	0.0728	0.0334	0.5	2655	0.4
45	0.0225	0.0368	0.5	3214	0.34
46	0.3486	0.0389	0.5	2282	3.8
47	0.0406	0.0361	0.5	1888	0.21

correlation between variable is both strongly correlated and statistically significant.

(5) Model training, validation and testing.

The normalised data in step (4) was randomly separated into training set and testing set by a ratio of 7:3. The XGBoost model was developed based on the training set, and then testing set was applied to test the prediction accuracy of this model. Due to lacking independent testing dataset, the original data was inherently resampled to evaluate the performance of model, which was also a widely used method when developing machine learning model. Additionally, the tuning parameters of XGBoost were optimised through GridSearch by using K-fold cross-validation (Novellino et al. 2021). The initial data was usually assigned to K partitions with equal size. Then, one of the partitions was selected randomly to be labeled as testing set and the other $K-1$ partitions were used for model training. Each of the partitions was labeled once to serve as testing set. As a result, the average result was exported as the final output after K times cross-validation. Finally, the optimal parameters were used to develop the best training model to generate prediction results. The overall regression mechanism of XGBoost was presented in the Appendix.

(6) Model assessment and validation.

The generation of prediction results using XGBoost method relies on the stability of the input data and also the hyper-parameters, such as 'learning_rate', 'max_depth', 'min_child_weight' and 'n_estimators', MAPE and R-squared were selected to evaluate the performance of this model. Meanwhile, the other three regression methods, RF, GBDT and AdaBoost, were introduced to conduct comparative analysis.

MAPE and R-squared are two widely used approaches to detect the errors between prediction and true values and evaluate the performance of model, in which MAPE can be calculated as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (2)$$

where $\hat{y}_i$ represents the prediction value and y_i is the true value. n indicates the number of prediction results. MAPE ranges from 0 to positive infinity and a better model is indicated if MAPE is closer to 0. Apart from the MAPE which reflects both the error between prediction and true values, R-squared was used to evaluate the training model. It can be demonstrated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\overline{y_i} - y_i\right)^2} \quad (3)$$

where $\hat{y}_i$ and y_i represent the prediction and true values, respectively. $\overline{y_i}$ is the mean value of all the true values. Therefore, R-squared reflects the error between the developed model and baseline model, which means that a better model is developed when R-squared is close 1 and even equal to 1. Finally, the prediction results of training model in step (5) were compared with other empirical equations and regression models.

Result analysis

Results of correlation analysis

As indicated in Fig. 6, the seven scatter diagrams demonstrate the correlations between the selected factors and landslide volume. Area of slope units (S) shows strong correlation with landslide volume (V) based on Person's coefficient (r) and Determination Coefficient (R^2). It means that the slope units with large area are more likely to result in large-scale landslide. Apart from the area of slope units, fault density (Fd), mean elevation of slope units (El) and geological index (GI) also present strong correlation with landslide volume (V) because r values of the three factors were all larger than 0.5. The p-values of all four factors are smaller than 0.01, which demonstrates the significance between these factors and landslide volume. However, the other three factors, mean NDVI, mean slope gradient of slope units (Sl) and mean curvature of slope units (Slc), show weak correlation with landslide volume since the r values of them were all smaller than 0.4. Therefore, $NDVI$, Sl and Slc were excluded from prediction of landslide volume. The four factors, S, Fd, El and GI, were finally decided for model training.

Establishment of dataset

As illustrated in the above section that S, Fd, El and GI were decided as input data for model traning and part of the collected data is presented in following Table 5.

Furthermore, the distribution frequency of landslides is presented in Fig. 7 based on the collected data, which demonstrated the landslide-inducing environment in this area. The same percentage of number of landslides was identified with the S ranging from 0.0355 to 0.0525 and 0.0695 to 0.0865 km², reaching 17%, while the high faults density failed to increase the occurrence frequency of landslide according to Fig. 7b. Instead, 26% of occurred landslide concentrated around to the Fd of 0.037. This is because these landslides were mostly located in the south part of this area with abundant rainfall which triggered the happening of landslide. Unlike the distribution characteristics of S and Fd, El showed a relatively even distribution that the mean elevation of 57% slope units was observed from 2040 to 3240 m. Nevertheless, typical concentrated distribution was found in Fig. 7d, which indicated the relatively

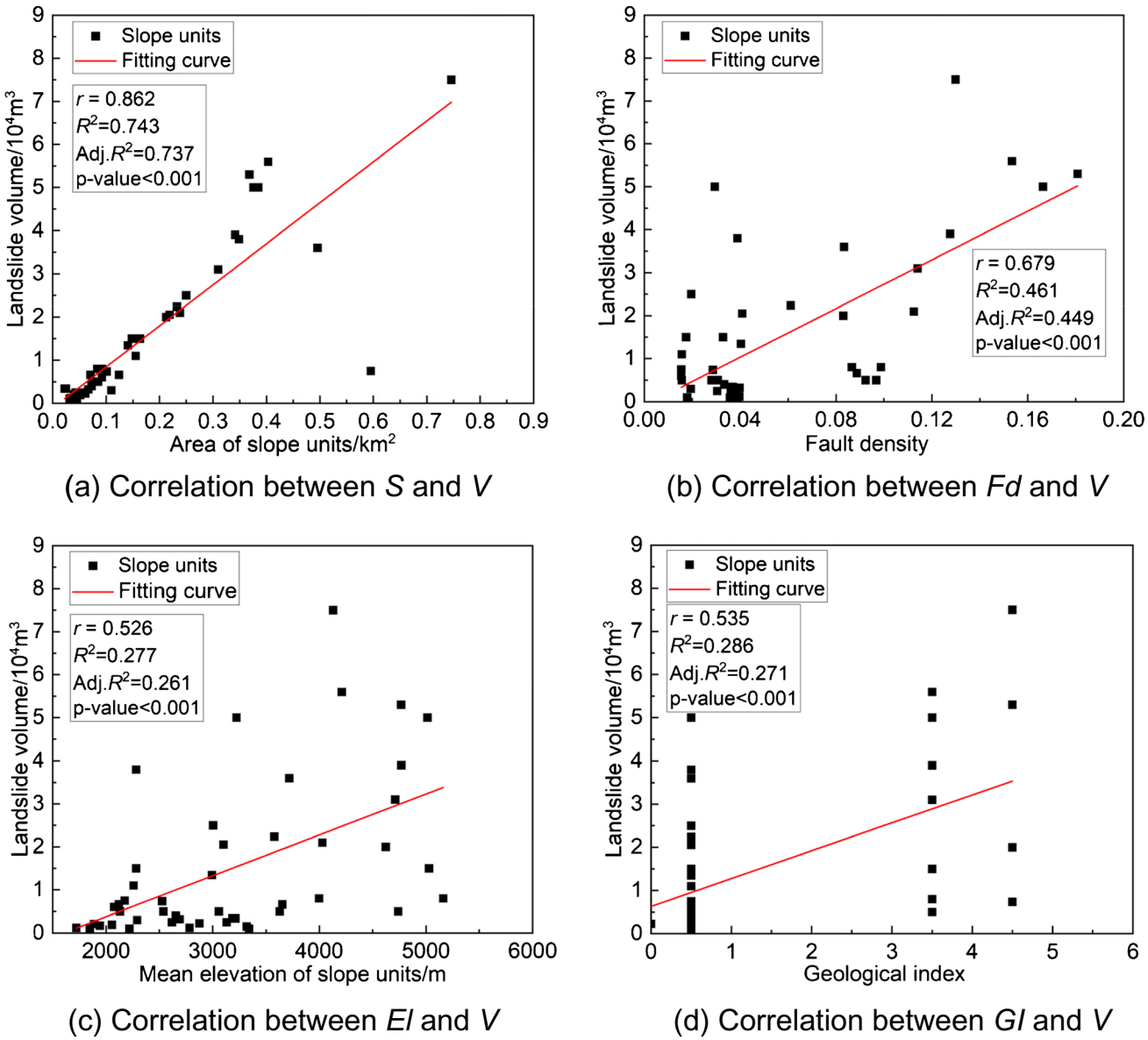
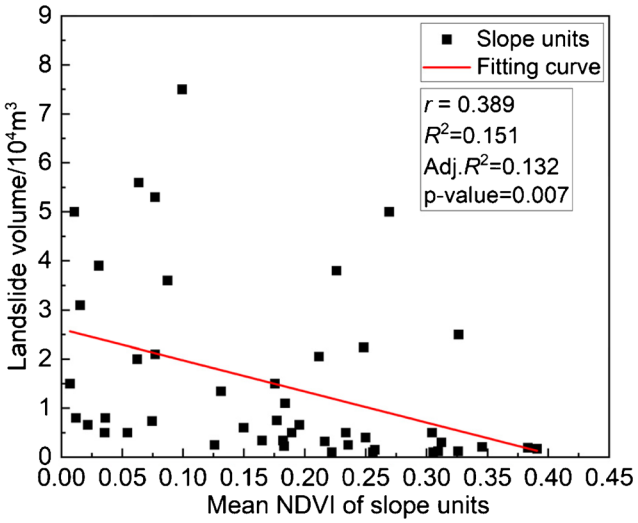


Fig. 6 Collinearity analysis between each controlling factor and landslide volume. ('S' represents area of slope units, 'Fd' represents the mean fault density of slope units, 'El' represents mean elevation of slope units, 'GI' is the geological index, 'NDVI' is normalised density vegetation index, 'Sl' is mean slope gradient of slope units, 'Slc' represents the mean curvature of slope units)

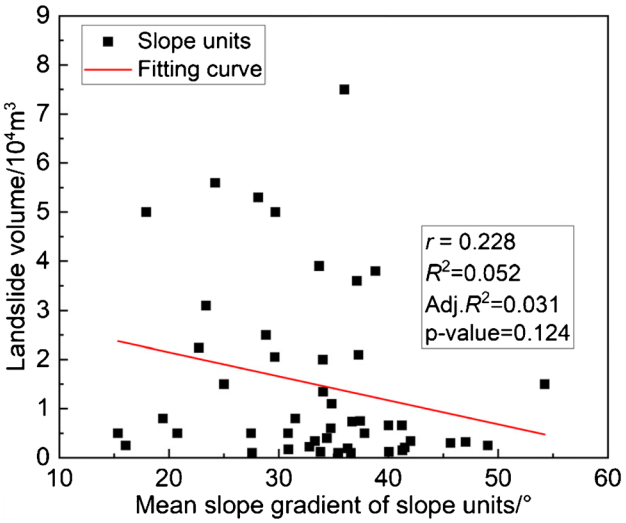
high erodibility of the rocks in this area because of the effect of water erosion and faults movement.

Prediction results of machine learning

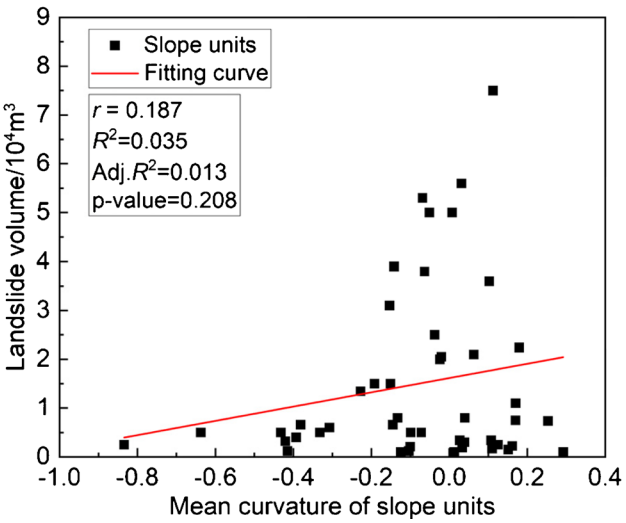
The four models, XGBoost, GBDT, AdaBoost and RF, were developed using Python within the PyCharm editing environment, based on the selected four controlling factors. The optimisation of XGBoost's hyperparameters was performed through a grid search with five-fold cross-validation, enabling the determination of optimal values for *learning_rate*, *max_depth* and *n_estimators*. Meanwhile, the same method was used to decide the optimal parameters of the other three models. The evaluation results are presented in Fig. 8. XGBoost achieves the smallest MAPE, 8.19%, and highest *R*-squared, 0.986, followed by GBDT which achieves MAPE of 16.87% and *R*-squared of 0.931. Moreover, RF ranks the third place when evaluating the performance of landslide volume prediction. The MAPE is found to be 24.38% and the *R*-squared is 0.907. There is no significant difference observed in R-squared for the three models, which means that the three models all perform well in model training. But the MAPE values of GBDT and RF indicate the poor generalisation ability of the two models when compared with XGBoost. Finally, AdaBoost achieves the highest MAPE and lowest *R*-squared, 27.6% and 0.764, respectively. Therefore, the XGBoost method proves its superiority in accuracy improvement, avoiding overfitting and enhancement of the generalisation ability



(e) Correlation between *NDVI* and *V*



(f) Correlation between *SI* and *V*



(g) Correlation between *Slc* and *V*

Fig. 6 (continued)

when compared with other boosting-related and bagging-related algorithms.

Further investigations are considered here to reflect the contributions of these factors in improving prediction accuracy and model development. As indicated in MAPE of Fig. 9, significant decline of MAPE can be achieved after the integration of fault density (*Fd*), lowering to 13.86%. Another remarkable reduction, 5.67%, is observed after the involvement of mean elevation of slope units (*El*) and geological index (*GI*), finally achieving a MAPE of 8.19%. As for the change of *R*-squared, distinct rise of 0.155 is observed when integrating both area of slope units (*S*) and fault density (*Fd*) into model training. Then, slight increase is achieved after the mean elevation of slope units (*El*) and geological index (*GI*) are involved into XGBoost, ultimately reaching 0.986. Moreover, in order to present the superiority of machine learning method in predicting landslide volume of rock slopes and selection of indicators in this paper, several empirical methods and regression model are introduced in next section to conduct comparison analysis between these methods.

Prediction comparisons with other empirical and regression methods

Numerous studies have examined the relationship between landslide area (A_L) and landslide volume (V_L) using historical landslide data, leading to the development of various semi-empirical equations for estimating potential landslide volume. In this section, several empirical equations (Table 6) are introduced to further assess the model's performance in this study, considering the selected factors in relation to geomorphic and geological conditions.

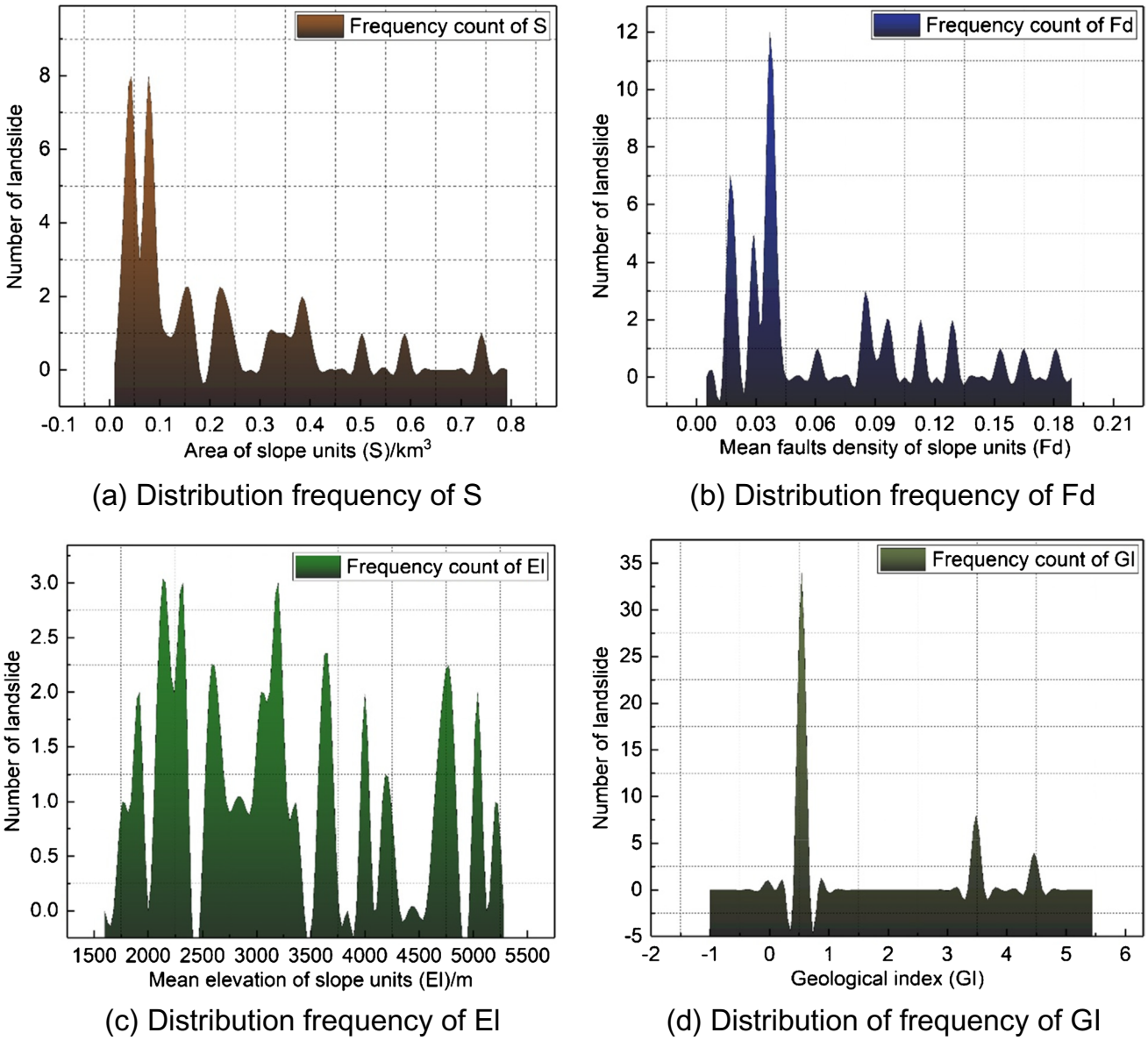


Fig. 7 Frequency count of selected factors

The comparison results between these methods are showed in Fig. 10. A slight overestimation is observed in the fitting curve of the predicted landslide volume compared to the measured landslide volume. However, among the methods evaluated, the machine learning approach demonstrates the best performance, as its regression results exhibit the lowest error relative to the measured landslide volume. In the case of a given rock slope, landslides can be triggered by factors such as earthquakes, rainfall and snowmelt (Guzzetti et al. 2009). The potential landslide volume is primarily influenced by geomorphic and geological conditions, which determine rock detachment and fracturing. While the affected area provides an indication of the surface extent of landslides, incorporating additional geological factors can further enhance prediction accuracy.

Discussion and limitations

The machine learning model has been demonstrated to outperform existing empirical formulas in predicting landslide volume for rock slopes. This improvement is attributed to the integration of a greater number of controlling factors into the model, without relying on simplifying assumptions in the analytical process. As indicated in the above analysis, the landslide area is widely used as indicator to estimate the landslide magnitude, but the lithology and orientation of rocks cannot be ignored if accurate landslide volume is expected because the understanding of the lithology and orientation of rocks can benefit the determination of the possible thickness of slide mass. Additionally, geological activities, such as fault movements, further weaken the internal structure of rock slopes, indirectly influencing the volume of the slide mass. More critically,

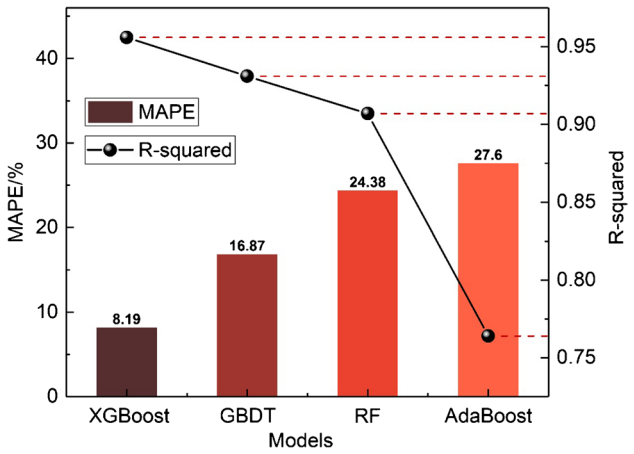


Fig. 8 Performance of different ML models in predicting landslide volume

this study area is located in the Middle Himalayas, a region characterised by frequent geotectonic activity, which further exacerbates slope instability. As a consequence, landslide frequently occurred with magnitude mainly smaller than 10^4 m^3. Therefore, the limitation of landslide volume prediction in this paper may appear that this developed model can achieve high accuracy in predicting the landslide magnitude ranging from 10^3 to 10^4 m^3, while the volume prediction of medium (10^5–10^6) and large size (10^6–10^7) landslide needs further investigations. Moreover, although an optimal combination of factors was proposed in this study to enhance volume prediction, other factors not included in the model, such as rainfall and seismic intensity, can also influence landslide magnitude in individual events. Despite certain limitations of the machine learning approach, the developed model significantly improves prediction accuracy and facilitates timely warnings for local communities. Furthermore, this research highlights the effectiveness of machine learning as an initial estimation tool for assessing potential landslide volume in rock slopes.

Conclusion

In conclusion, this study evaluates the effectiveness of machine learning in predicting landslide volume based on the proposed combination of controlling factors for rock slopes in Gyirong, China. The prediction is carried out through the following key steps: (1) identifying historical landslides from the past five years using remote sensing imagery and field investigations, (2) selecting controlling factors and constructing a dataset, (3) performing collinearity analysis to assess the correlations between each factor and landslide volume, (4) developing the model using normalised data, followed by testing and validation to evaluate its performance, and (5) comparing the machine learning model with empirical equations to assess its predictive accuracy.

A total of 82 landslides were identified in this area, of which 47 were used for landslide volume prediction. The remaining landslides were excluded due to the inability to estimate their volume, as plant growth and long-term water erosion have obscured their original features. Based on the landslide identification results, seven factors related to geomorphic, geological and ecological conditions were selected for landslide volume estimation. However, collinearity analysis using Pearson's coefficient (r) and p-value reveals that only four

Table 6 Empirical methods in the past studies

No	Equation	Source
1	$V_L = 0.074 \times A_L^{1.45}$	Guzzetti et al. (2009)
2	$V_L = 0.19 \times A_L^{1.19}$	Imaizumi et al. (2008)
3	$V_L = 0.39 \times A_L^{1.31}$	Imaizumi and Sidle (2007)

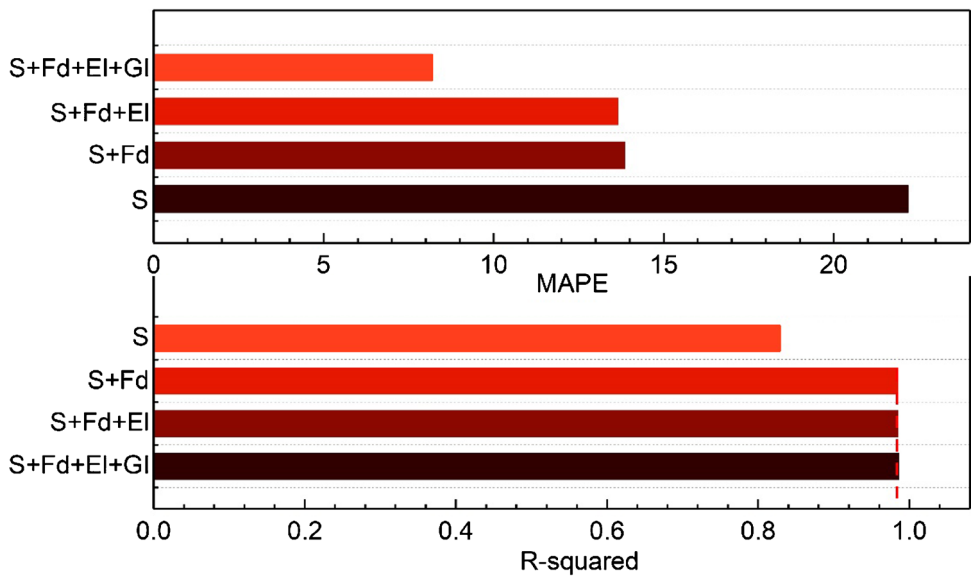


Fig. 9 Factors contributions in improvement of volume prediction

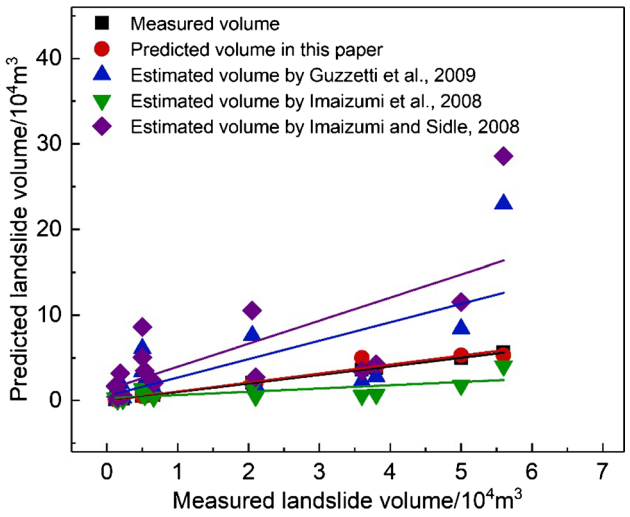


Fig. 10 Comparison of volume estimation of landslide for rock slopes by different methods

of these factors exhibit a strong correlation with landslide volume, including area of slope units (*S*), mean elevation of slope units (*El*), faults density (*Fd*) and geological index (*GI*). After that, the extracted data is divided into a training set and a testing set to develop the XGBoost, GBDT, AdaBoost and RF models. The testing set is then used to evaluate the models' predictive performance based on Mean Absolute Percentage Error (MAPE) and R-squared (R^2) values. The results demonstrate that XGBoost achieves the highest R^2 value (0.986) and the lowest MAPE (8.19%), indicating its superior performance. Furthermore, several empirical equations are introduced to assess the effectiveness of the XGBoost model, and the results confirm that it achieves the lowest prediction error. Lastly, the distribution characteristics of landslides and litho-stratigraphic classes are analysed to examine the influence of lithology and rock further enhance prediction accuracy.

Acknowledgements

This study was financially supported by the Suqian Sci & Tech program [Grant No. S202218]; Jiangsu Province Industry-University-Research Cooperation Project, China [Grant No. BY20231250].

Data Availability

Data will be made available on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Appendix

The regression mechanism of XGBoost was described as:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_x(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{4}$$

where $\hat{y}_i^{(t)}$ represents the prediction result of sample i after t_{th} iteration, and $\hat{y}_i^{(t-1)}$ is the addition of the prediction result for all the (t-1) decision trees. $f_t(x_i)$ represents the function of the t_{th} tree. In order to obtain the overall model, the function of each iteration should be developed firstly and then the overall model can be produced through the addition of all the functions. Therefore, first of all, the loss function can be demonstrated by:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \tag{5}$$

where n is the quantity of samples. $\hat{y}_i$ and y_i represent prediction value and true value respectively. The prediction accuracy of model is decided based on the deviation value and variance value, in which the former is represented by the loss function. Meanwhile, the variance value should be as small as possible to ensure the prediction accuracy. Therefore, the regularisation term is added to the loss function in order to avoid overfitting and improve prediction accuracy:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{6}$$

where Obj represents the target function, $\sum_{i=1}^{t} \Omega(f_i)$ is the regularisation term. Substituting Eq. (2) into Eq. (4):

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \sum_{i=1}^{t-1} \Omega(f_i) \tag{7}$$

where $\sum_{i=1}^{t-1} \Omega(f_i)$ can be regarded as a constant since all of the (t-1) trees have been decided. The only unknown term in this equation is $\Omega(f_t)$. Finally, Taylor's expansion is introduced to Eq. (7):

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + \sum_{i=1}^{t-1} \Omega(f_i) \tag{8}$$

where $Obj^{(t)}$ represents the overall model.

References

Ali N, Farshchi I, Mu'azu MA et al (2012) Soil-root interaction and effects on slope stability analysis. Electron J Geotech Eng 17:319–328

Alvioli M, Guzzetti F, Rossi M (2014) Scaling properties of rainfall induced landslides predicted by a physically based model. Geomorphology 213:38–47

Apip ST, Tachikawa Y, Takara K, Yamashiki Y (2010) Assessing sources of parametric uncertainty and uncertainty propagation in sediment runoff simulations of flooding. J Flood Risk Manag 3:270–84

Arabameri A, Saha S, Roy J et al (2020) Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash river watershed, Iran. Remote Sensing 12(3):475

Blahut J, van Westen CJ, Sterlacchini S (2010) Analysis of landslide inventories for accurate prediction of debris-flow source areas. Geomorphology 119(1–2):36–51

Budimir MEA, Atkinson PM, Lewis HG (2015) A systematic review of landslide probability mapping using logistic regression. Landslides 12(3):419–436

Chen W, Li Y (2020) GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. CATENA 195:104777. https://doi.org/10.1016/j.catena.2020.104777

D' Agostino V, Marchi L (2001) Debris flow magnitude in the Eastern Italian Alps: data collection and analysis. Phys Chem Earth C 26(9):657–663

Di Napoli M, Carotenuto F, Cevasco A et al (2020) Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. Landslides 17(8):1897–1914

Gaziev E (2013) Rock slope stability assessment in the fernando hiriart arch dam reservoir. In: Margottini C, Canuti P, Sassa K (eds) Landslide science and practice. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-31319-6_5

Gerrard AJ (1990) Mountain environments: an examination of the physical geography of mountains. Belhaven Press, London

Giles PT, Franklin SE (1998) An automated approach to the classification of the slope units using digital data. Geomorphology 21(3-4):251–264

Guzzetti F, Ardizzone F, Cardinali M et al (2009) Landslide volume and landslide mobilization rates in Umbria, Central Italy. Earth Planet Sci Lett 279(3–4):222–229

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31(1-4):181–216

Hovius N, Stark CP, Allen PA (1997) Sediment flux from a mountain belt derived by landslide mapping. Geology 25:231–234

Imaizumi F, Sidle RC (2007) Linkage of sediment supply and transport processes in Miyagawa Dam catchment, Japan. J Geophys Res: Earth Surface 112(F3). https://doi.org/10.1029/2006JF000495

Imaizumi F, Sidle RC, Kamei R (2008) Effects of forest harvesting on the occurrence of landslides and debris flows in steep terrain of central Japan. Earth Surface Processes and Landforms: the Journal of the British Geomorphological Research Group 33(6):827–840

IPCC (2013) Climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Cambridge University Press, Cambridge, United Kingdom and New York, p 1535

Jomelli V, Pavlova I, Eckert N et al (2015) A new hierarchical Bayesian approach to analyse environmental and climatic influences on debris flow occurrence. Geomorphology 250:407–421

Kadavi PR, Lee CW, Lee S (2018) Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sensing 10(8):1252

Kirschbaum D, Adler R (2013) Evaluation of landslide inventory information: extreme precipitation and global patterns. In: Margottini C, Canuti P, Sassa K (eds) landslide science and practice. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31313-4_17

Lacasse S, Nadim F, Lacasse S, Nadim F (2009) Landslide Risk assessment and mitigation strategy. In: Sassa K, Canuti P (eds) Landslides – Disaster Risk Reduction. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69970-5_3

Lee DH, Cheon E, Lim HH et al (2021) An artificial neural network model to predict debris-flow volumes caused by extreme rainfall in the central region of South Korea. Eng Geol 281:105979

Liao Z, Hong Y, Kirschbaum D, Adler RF, Gourley JJ, Wooten R (2011) Evaluation of TRIGRS (transient rainfall infiltration and grid-based regional slope-stability analysis)'s predictive skill for hurricane-triggered landslides: a case study in Macon County, North Carolina. Nat Hazards 58:325–339

Marjanović M, Kovačević M, Bajat B et al (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Eng Geol 123(3):225–234

Martinello C, Cappadonia C, Conoscenti C et al (2021) Optimal slope units partitioning in landslide susceptibility mapping. J Maps 17(3):152–162

Muntohar AS, Liao HJ (2009) Analysis of rainfall-induced infinite slope failure during typhoon using a hydrological–geotechnical model. Environ Geol 56:1145–1159

Novellino A, Cesarano M, Cappelletti P, Di Martire D, Di Napoli M, Ramondini M, Sowter A, Calcaterra D (2021) Slow-moving landslide risk assessment combining Machine Learning and InSAR techniques. Catena 203:105317

O'Gorman PA (2015) Precipitation extremes under climate change. Current Climate Change Reports 1(2):49–59

Perov V, Chernomorets S, Budarine O et al (2017) Debris flow hazards for mountain regions of Russia: regional features and key events. Nat Hazards 88(1):199–235

Petrakov DA, Krylenko IV, Chernomorets SS, Tutubalina OV, Krylenko IN, Shakhmina MS (2007) Debris flow hazard of glacial lakes in the Central Caucasus. In: 4th Int. conf. on debris-flow hazards mitigation, Millpress, Rotterdam, pp. 703–714

Piteau DR, Peckover FL (1978) Engineering of rock slopes. Landslides Analysis and Control 176:192–228

Pourghasemi HR, Gayen A, Park S et al (2018) Assessment of landslide-prone areas and their zonation using logistic regression, logitboost, and naïvebayes machine-learning algorithms. Sustainability 10(10):3697

Qiu C, Su L, Zou Q, Geng X (2022) A hybrid machine-learning model to map glacier-related debris flow susceptibility along Gyirong Zangbo watershed under the changing climate. Sci Total Environ 818:151752

Qiu C, Su L, Bian C, Zhao B, Geng X (2024) An AI-based method for estimating the potential runout distance of post-seismic debris flows. Int J Disaster Risk Sci 15(4):608–621

Qiu H, Cui P, Regmi AD, Hu S, Wang X, Zhang Y, He Y (2017) Influence of topography and volume on mobility of loess slides within different slip surfaces. Catena 157:180–188

Rahmati O, Tahmasebipour N, Haghizadeh A et al (2017) Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298:118–137

Simonett DS (1967) Landslide distribution and earthquakes in the Bewani and Torricelli Mountains, New Guinea. In: University C (ed) Landform Studies from Australia and New Guinea, Jennings JN, Mabbutt JA. Press, Cambridge, pp 64–84

Singhroy V (2009) Satellite remote sensing applications for landslide detection and monitoring. In: Sassa K, Canuti P (eds) Landslides – Disaster Risk Reduction. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-69970-5_7

Stanley T, Kirschbaum DB (2017) A heuristic approach to global landslide susceptibility mapping. Nat Hazards 87(1):145–164

Sun X, Chen J, Han X, Bao Y, Zhan J, Peng W (2020) Application of a GIS-based slope unit method for landslide susceptibility mapping along

the rapidly uplifting section of the upper Jinsha River, South-Western China. Bull Eng Geol Environ 79:533–549
Tanyaş H, Van Westen CJ, Allstadt KE et al (2017) Presentation and analysis of a worldwide database of earthquake-induced landslide inventories. J Geophys Res Earth Surf 122(10):1991–2015
Tanyaş H, van Westen CJ, Persello C et al (2019) Rapid prediction of the magnitude scale of landslide events triggered by an earthquake. Landslides 16(4):661–676
Tien Bui D, Tuan TA, Klempe H et al (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2):361–378
Wang M, Shen ZK (2020) Present-day crustal deformation of continental China derived from GPS and its tectonic implications. Journal of Geophysical Research: Solid Earth 125(2):e2019JB018774
Xie M, Esaki T, Zhou G (2004) GIS-based probabilistic mapping of landslide hazard using a three-dimensional deterministic model. Nat Hazards 33:265–282
Youssef AM, Pourghasemi HR, Pourtaghi ZS et al (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13:839–856
Zhang Y, Meng XM, Dijkstra TA, Jordan CJ, Chen G, Zeng RQ, Novellino A (2020) Forecasting the magnitude of potential landslides based on InSAR techniques. Remote Sens Environ 241:111738
Zhao B, Su L, Qiu C, Lu H, Zhang B, Zhang J, Geng X, Chen H, Wang Y (2024) Understanding of landslides induced by 2022 Luding earthquake, China. J Rock Mech Geotech Eng. https://doi.org/10.1016/j.jrmge.2024.07.006
Zhou C, Yin K, Cao Y et al (2018) Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. Comput Geosci 112:23–37

Chongcai Xu
School of Information Engineering, Suqian University, Suqian 223800, China

Congchao Bian
School of Information Science and Engineering, Hohai University, Nanjing 210098, China

Teng Yu
School of Civil Engineering and Architecture, Suqian University, Suqian 223800, China

Chenchen Qiu (✉)
Civil and Environmental Engineering, University of Strathclyde, Glasgow G1 1XQ, UK
Email: chenchen.qiu@strath.ac.uk