frontiers | Research Topics

Geo-information for geohazard and georisk

Edited by
Lixin Wu, Wei Zhao, Zefa Yang and Arun Saraf

Published in
Frontiers in Earth Science
Frontiers in Environmental Science

ISSN 1664-8714
ISBN 978-2-8325-3515-8
DOI 10.3389/978-2-8325-3515-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Geo-information for geohazard and georisk

Topic editors

Lixin Wu — Central South University, China

Wei Zhao — Institute of Mountain Hazards and Environment, Chinese Academy of Sciences (CAS), China

Zefa Yang — Central South University, China

Arun Saraf — Indian Institute of Technology Roorkee, India

Citation

Wu, L., Zhao, W., Yang, Z., Saraf, A., eds. (2023). *Geo-information for geohazard and georisk*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3515-8

Table of contents

frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 12 October 2022
DOI 10.3389/fenvs.2022.1027423

OPEN ACCESS

EDITED BY
Zefa Yang,
Central South University, China

REVIEWED BY
Quoc Bao Pham,
University of Silesia in Katowice, Poland
Osman Orhan,
Mersin University, Turkey

*CORRESPONDENCE
Aqil Tariq,
at2139@msstate.edu,
aqiltariq@whu.edu.cn
Linlin Lu,
lull@radi.ac.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing,
a section of the journal
Frontiers in Environmental Science

RECEIVED 25 August 2022
ACCEPTED 23 September 2022
PUBLISHED 12 October 2022

CITATION
Islam F, Riaz S, Ghaffar B, Tariq A, Shah SU, Nawaz M, Hussain ML, Amin NU, Li Q, Lu L, Shah M and Aslam M (2022), Landslide susceptibility mapping (LSM) of Swat District, Hindu Kush Himalayan region of Pakistan, using GIS-based bivariate modeling.
Front. Environ. Sci. 10:1027423.
doi: 10.3389/fenvs.2022.1027423

Landslide susceptibility mapping (LSM) of Swat District, Hindu Kush Himalayan region of Pakistan, using GIS-based bivariate modeling

Fakhrul Islam[1], Salma Riaz[2], Bushra Ghaffar[3], Aqil Tariq[4,5]*, Safeer Ullah Shah[6], Muhammad Nawaz[7], Mian Luqman Hussain[8], Naz Ul Amin[9], Qingting Li[10], Linlin Lu[11]*, Munawar Shah[12] and Muhammad Aslam[13]

[1]Department of Geology, Khushal Khan Khattak University, Karak, Pakistan, [2]Department of Applied Mathematics and Statistics, Institute of Space Technology, Islamabad, Pakistan, [3]Department of Environmental Science, International Islamic University Islamabad, Islamabad, Pakistan, [4]Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Starkville, MS, United States, [5]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, [6]Ministry of Climate Change, Islamabad, Government of Pakistan, Islamabad, Pakistan, [7]International Centre for Integrated Mountain Development, Islamabad, Pakistan, [8]National Centre of Excellence in Geology, Peshawar, Pakistan, [9]Pakistan Forest Institute, Peshawar, Pakistan, [10]Airborne Remote Sensing Center, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [11]Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [12]Department of Space Sciences, Space Education and GNSS Lab, National Center of GIS and Space Application, Institute of Space Technology, IST, Islamabad, Pakistan, [13]School of Computing Engineering and Physical Sciences, University of West of Scotland, Paisley, United Kingdom

Landslides are a recurrent environmental hazard in hilly regions and affect the socioeconomic development in Pakistan. The current study area is the tourism and hydro energy hub of Pakistan and is affected by environmental hazard. A landslide susceptibility mapping (LSM) of the Hindu Kush Himalayan, Swat District, Pakistan, can be created to reduce demographic losses due to landslides. This current study is conducted to apply three bivariate models, including weights of evidence (WOE), frequency ratio (FR), and information value (IV) for an LSM that has not been explored or applied in the current study area. For this purpose, first, an inventory map of 495 landslides was constructed from both ground and satellite data and randomly divided into training (70%) and testing (30%) datasets. Furthermore, 10 conditioning factors (elevation, slope, aspect, curvature, fault, rainfall, land use land cover (LULC), lithology, road, and drainage) used for the mapping of landslides were prepared in ArcGIS 10.8. Finally, LSM is generated based on WOE, FR, and IV models and validated the performance of LSM models using the area under receiver operating characteristic curve (AUROC). The findings of success rate curve (SRC) of the WOE, FR, and IV models were 67%, 93%, and 64%, respectively, while the prediction rate curves (PRCs) of the three models were 87%, 95%, and 73%, respectively. The validation results for WOE, FR, and IV justified that the FR model is the most reliable technique of all three of these models to produce the

highest accuracy LSM for the present study area. Policymakers can use the findings of current research work to mitigate the loss due to landslide hazard.

KEYWORDS

landslide susceptibility mapping, landslide causative factors, geospatial modeling, bivariate models, sub Himalayas

1 Introduction

Landslides triggered in mountainous areas can have catastrophic consequences, threaten human life, and cause billions of dollars in economic losses. A landslide is a geological process in which material moves downward from its parent rocks due to slope failure (Zhang et al., 2020; Zhou W. et al., 2021; Yue et al., 2021; Zhu B. et al., 2022). Landslides are the most catastrophic geo-environmental hazard in mountainous regions, severely impacting the world's socioeconomic trends (Ullah et al., 2022). Landslide hazard is a frequent disaster phenomenon in the mountainous regions in developing countries like Pakistan (Zhang et al., 2019b, 2019a; Wang S. et al., 2021; Quan et al., 2021). Landslide hazard is expected to grow due to anthropogenic interventions such as deforestation, population growth, urbanization, etc. and natural phenomena due to topographic, geologic, and climatic factors (Rahman et al., 2020; Xie et al., 2021a, 2021b; Guo et al., 2022; Ren et al., 2022; Wahla et al., 2022). Since the 20th century, mortality caused landslides has reached 6.2 million and has caused up to 10 billion US dollars of damage. It is essential to assess the anthropogenic and natural predisposing factors of landslides to reduce their consequences in mountainous areas (Raghuvanshi, 2019; Tariq et al., 2021b; Yin et al., 2022a, 2022b; Zhu Z. et al., 2022; Chen et al., 2022).

It is expected that landslides will originate in certain areas due to the combined effects of many causative parameters instead of a single causative factor. Therefore, assessing all-natural and human-induced parameters and their association with landslide inventory can help develop LSM for predicting landslides to minimize the damages of said hazard (Girma et al., 2015; Chimidi et al., 2017; Zhang et al., 2021; Shah et al., 2022). An LSM may be produced based on the postulation that the predisposing factors for the imminent landslide will be the same as in earlier landslides (Pham et al., 2015; Zhou et al., 2021a, 2021b; Sharifi et al., 2022; Zhan et al., 2022). An LSM demarcates the region exposed to landslide hazards and classifies the area into different potential hazard zones by using various approaches to manage the landslide hazard in the region of interest. Researchers have developed various GIS-based techniques for landslide modeling (Pourghasemi et al., 2005; Reichenbach et al., 2018; Tian et al., 2021a; Wang P. et al., 2021, 2022; Fu et al., 2022).

Generally, these methods can be grouped into qualitative and quantitative versions to construct detailed LSM of the area. The qualitative approach is subjectively used by researchers based on their relevant experience and judgment. A quantitative method is an objective-based method in which the inventory of landslides and causative factors is evaluated through mathematical and statistical ways to produce LSM (Milevski et al., 2019; Majeed et al., 2022). Qualitative and quantitative methods can be used to assess landslide inventories and causative factors, and both techniques incorporate various pros and cons in the research (Tian et al., 2019, 2020; Tian et al., 2021b; Chen et al., 2021; Li Y. et al., 2022). Qualitative methods play significant roles in LSMs, but this technique is time-consuming in landslide research (Barredo et al., 2000; Zuhairi et al., 2020; Zhao et al., 2021; Li Q. et al., 2022; Sadiq Fareed et al., 2022). An LSM is generated using quantitative methods like bivariate and multivariate analysis in the modern era due to their high accuracy results and easy implementation. These approaches can compute the impacts of each class of causative factor with landslide events (Lee and Pradhan, 2007; Dahal et al., 2008; Pradhan, 2010; Choi et al., 2012; Park et al., 2013; Vakhshoori and Zare, 2016; Fayez et al., 2018). The GIS-based statistical approach used the association of landslide inventory and predisposing landslide factors to generate LSM (Khanchou et al., 2020; Narimah Samat and Ismail, 2020). This study we used bivariate statistical models, such as weights of evidence (WOE), frequency ratio (FR), and information value (IV), to compute the association between landslide events and causative factors.

In this study, we used bivariate statistical models like WOE, FR, and IV to compute the mathematical association between landslide events and causative factors. These quantitative models quantify the authentic and reliable association between the variables more accurately than in the qualitative and semi-quantitative models. WOE is a GIS-based statistical model that estimates probability using the Bayesian principle and the concepts of prior and posterior probability. This technique is mostly applied in prediction mapping (Elmoulat et al., 2015; Hussain et al., 2022). The FR technique is considered to be a consistent and experimental method that can be used to calculate the association of landslide events with causative factors and to produce applicable LSM of the study area (Oh et al., 2017; Farhan et al., 2022). This model (IV) is a GIS-based bivariate models and has been applied to produce predicted study maps based on dependent and independent variables (Li et al., 2021; Tariq et al., 2021c; Imran et al., 2022).

Swat is Pakistan's emerging tourism and hydel energy hub and has been affected by terrorism and various natural hazards, i.e., floods, landslides, etc. Many researchers have worked on different hazards using diverse techniques and models. Some researchers have focused on floods, while few have concentrated

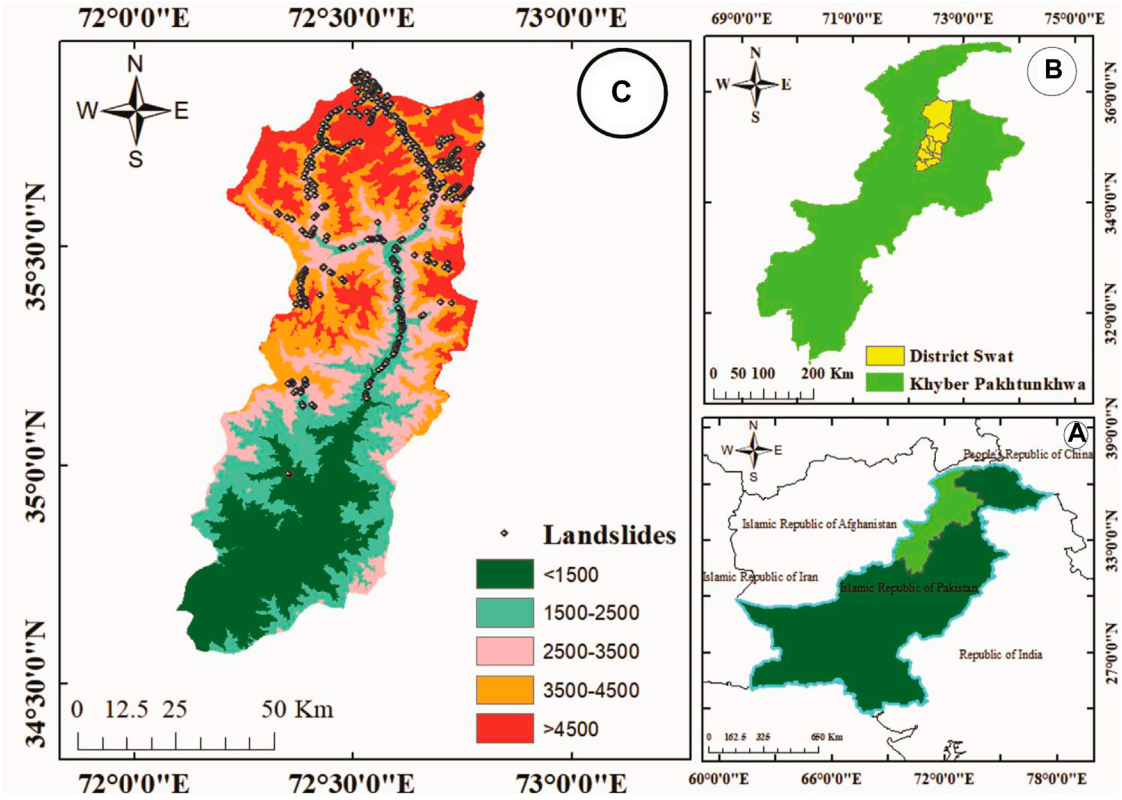


FIGURE 1
Map of the study area along with elevation in meters and spatial location of landslide events: **(A)** Pakistan administrative boundaries of KPK province; **(B)** Swat District in KPK shapefile; **(C)** elevation and landslide inventory of study area.

on slope instability in the Swat region. However, scientists of various fields ignored landslide investigation in the current study area using satellite data, Google Earth Engine (GEE), and geospatial modeling. In this study, we used GIS-RS-based emerging and innovative techniques to fill the research gap in the Hindu Kush Himalaya Ranges in Swat District of Pakistan. We selected this area because it has not been explored and nor investigated by researchers regarding landslide mapping using integrated study of GIS-RS with ground truth information.

The main goal of the present research work is to assess the quantities association of dependent variable (landslide events) with independent variables (causative factors) to generate an accurate, updated LSM of the study area using GIS-RS based methods and precariously emphasize the low, medium, high, and very high zone of landslides in Swat District of Pakistan. Finally, the LSM of models is validated by area under receiver operating characteristic curve (AUROC) technique and by ground data from the field, which can be helpful to decision makers at organizations to reduce the loss to hazard in the study region.

2 Materials and methods

2.1 Study area

This study is conducted in Swat District of Khyber Pakhtunkhwa (KPK), as shown in Figure 1. Geographically, the Swat District is situated within a latitude and longitude range of 34° 30′ 00″ 34° 30′ 00″’ North and 72° 05′ 00″’ and 72° 50′ 00″’ East, at 500–6,500 m elevation above sea level (Atta-ur-Rahman and Khan, 2011). Topographically, the study area is part of the Hindu Kush Himalayan range (Qasim and Hubacek, 2013), and the tourism hub of KPK has an area of 5,337 km^2. The study area is divided into semi-arid, sub-humid, and humid climate regions (Bahadar et al., 2015; Sharifi et al., 2022; Zamani et al., 2022). Geologically the research area is engraved in a Suture Zone (SZ) between the Indian Plate and Kohistan Island Arc (KIA). The SZ between KIA and the Asian Plate occurs on the Northside and is tectonically the most active geomorphic region (Tahirkheli, 1979; Abbas et al., 2021; Hu et al., 2021; Waqas et al., 2021).

TABLE 1 Detailed RS and ground data availability, statement, source, and purpose.

Data	Availability of data	Data availability statement/Source	Parameter maps
Landsat-8	Data openly available	The data supporting this study's findings are openly available in [USGS] at https://earthexplorer.usgs.gov. The spatial resolution is 30 m	Land use land cover
Sentinel-2	Data openly available	The data supporting this study's findings are openly available in [ESA] at https://www.esa.int. The sentinel -2 data of 10 m of spatial resolution	Landslide inventory
DEM (ALOS)	Data openly available	The data supporting this study's findings are openly available at https://asf.alaska.edu/. Spatial resolution is 12.5 m	Elevation, slope, aspect, curvature, and drainage
CHIRPS	Data openly available	The data supporting this study's findings are available in [UCSB] at https://www.chc.ucsb.edu. The spatial resolution of CHIPRS is 0.05° (5.54 km) and daily gridded	Rainfall maps
Rainfall Surface Precipitation Gauge (SPG)	Available in PMD organization on through proper channel based	The SPG data for 2010 to 2020 were acquired from PMD Peshawar regional center. https://www.pmd.gov.pk/en/.	Rainfall maps
Geological data	From Northern Map of Pakistan	The tectonic evolution of the Kohistan-Karakoram collision belt along the Karakoram Highway transect, north Pakistan. *Tectonics* 1999, *18*, 929–949. Scale 1:650, 000	Lithological and tectonic map
Road	Available in PKHA organization on through proper channel based	https://pkha.gov.pk/ Scale 1:25000	Proximity to road

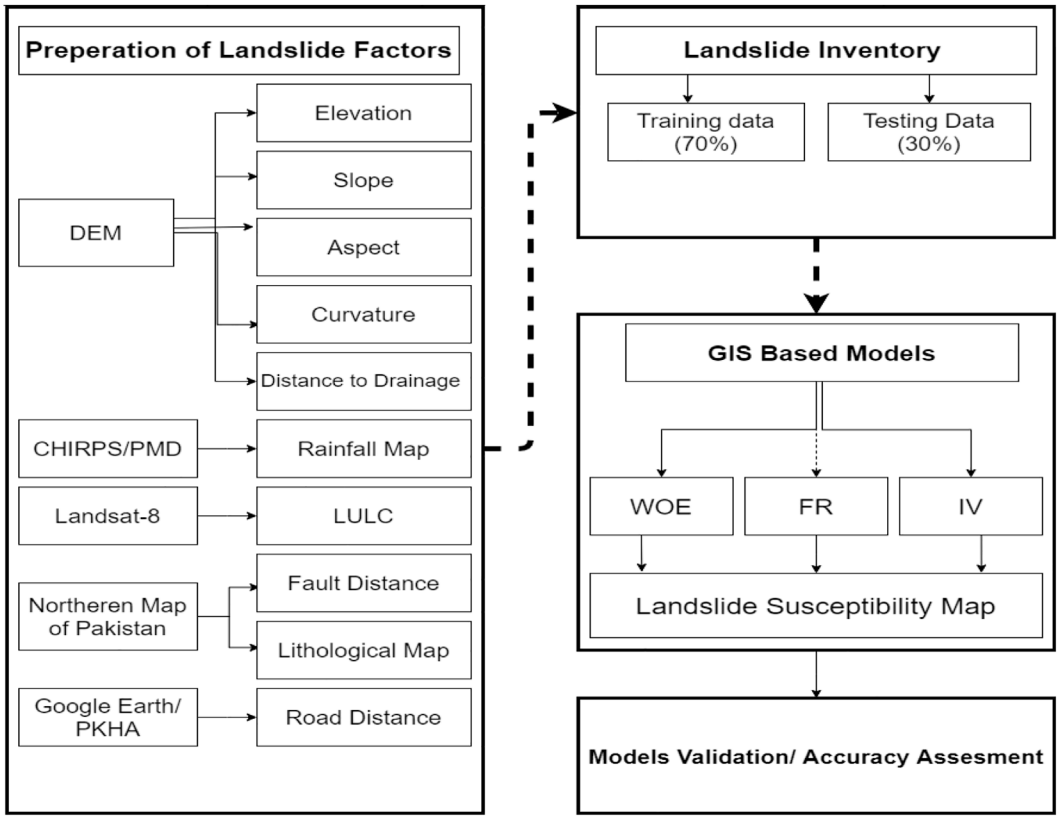


FIGURE 2
Methodology for GIS and RS-based landslide susceptibility mapping.

2.2 Datasets

In this research, we used both satellite and ground datasets to achieve the goal of the present study. The satellite data used in this study was composed of Landsat-8 Operational Land Imager (OLI), Sentinel-2, Advanced Land Observing Satellite Digital Elevation Model (ALOS DEM), Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), and Google Earth. Landsat-8 images from July 2020 with a spatial resolution of 30 m were used for the LULC classification of the present study area. Sentinel-2 data were used to detect and map landslide events. We used sentinel-2 images with Google Earth and a field survey to produce an updated landslide inventory map. ALOS DEM data were downloaded from Earth data, Alaska Satellite Facility (ASF), to compute the topographic causative factors. CHIRPS data from 2010 to 2020 were used to prepare a rainfall map of the study area.

The ground data consist of rainfall, geological, and road network data. The daily rainfall data from three meteorological stations, i.e., Saidu Sharif (34.73°N, 72.35°E), Kalam (35.52°N, 72.54°), and Malam Jabba (34.83°N, 72.55°E), for 2010–2020 were obtained from Pakistan Meteorological Department (PMD), Peshawar Regional Center (Tariq et al., 2022a; 2022b). The geological map was scanned and digitized from the geological map of northern Pakistan to prepare a lithological and tectonic map of the study area. The updated road data was acquired from KPK Highway Authority (PKHA). The details of both RS and ground datasets of the current study area are mentioned in Table 1. Further details are presented in Figure 2.

3 Methodology

3.1 Landslide inventory map

Past and current landslide inventory data are a significant factor in predicting landslide potential in a study area (Guzzetti et al., 1995). Therefore, a landslide inventory map is the first mandatory element for generating and compiling authentic LSM of a study area (Aslam et al., 2022a). The inventory map is a significant parameter for performing various quantitative analyses and validating the models' accuracy (Chalkias et al., 2014; Baloch et al., 2021; Baqa et al., 2021; Shah et al., 2021). The landslide inventory map for this study was developed using Sentinel-2 and Google Earth images and validated with field survey data. In the research area, 495 past and present landslide events were detected from satellite imageries and ground-based data, as shown in Figure 1. After landslide inventory generation, we developed the non-landslide area of the research region. First we extract the landslide polygon from the study area polygon to generate non-landslide area. Then we used ArcGIS tools to produce random points in the study area which have been considered as non-landslide area in the current research. During a field survey we validated the satellite-based inventory of landslide and observed the impacts of various type of landslide on forest, water quality, infrastructure, and human loss. Landslide hazard has had significant effects on the forest and environment of Swat District. This environmental hazard also affected the infrastructure and engineering structures. The impacts of the different type of landslide are shown in Figure 3.

3.2 Causative parameters

The incidence of landslides is predisposed by the combined effects of topographic, hydrological, and geological parameters (Costanzo et al., 2012; Marchesini et al., 2014; Tariq and Shu, 2020; Tariq et al., 2021a). Therefore, selecting causative factors is a significant step toward producing LSM. In this paper, We selected 10 precondition factors as independent variables for developing LSM for the region of interest. These factors are elevation, slope, aspect, curvature, precipitation, land use land cover (LULC), proximity to the fault, lithology, road proximity, and drainage network. In the following paragraphs, we briefly present the significance of each causative factor in the LSM. The details of causative factors are as follows.

3.2.1 Elevation

Elevation is an influential precondition for landslide incidence (Tosic et al., 2014). Altitude is used for landslide susceptibility mapping and investigation (Dou et al., 2015). Elevation was extracted from advanced land observation satellite (ALOS) with a 12.5 m spatial resolution and reclassified into different classes using the natural break algorithm in ArcGIS 10.8, as shown in Figure 4A.

3.2.2 Slope

The slope is a crucial causative factor in landslide investigation because it triggers the downslope movement of loose sediment material (Nolasco-Javier et al., 2015). The slope of the current study area was computed from ALOS DEM having 12.5 m spatial resolution. The calculated slope was then reclassified into five classes in ArcGIS 10.8, as shown in Figure 4B.

3.2.3 Aspect

Aspect also plays an influential role in the instability of strata because it controls moisture in the rocks and soil due to wind and sunlight exposure in the study area. The slope aspect map in the current study area was calculated from ALOS DEM using ArcGIS 10.8 and reclassified into nine classes, as shown in Figure 4C.

3.2.4 Curvature

The curvature is a morphological parameter and has a significant role in LSM. The geomorphic structures of the

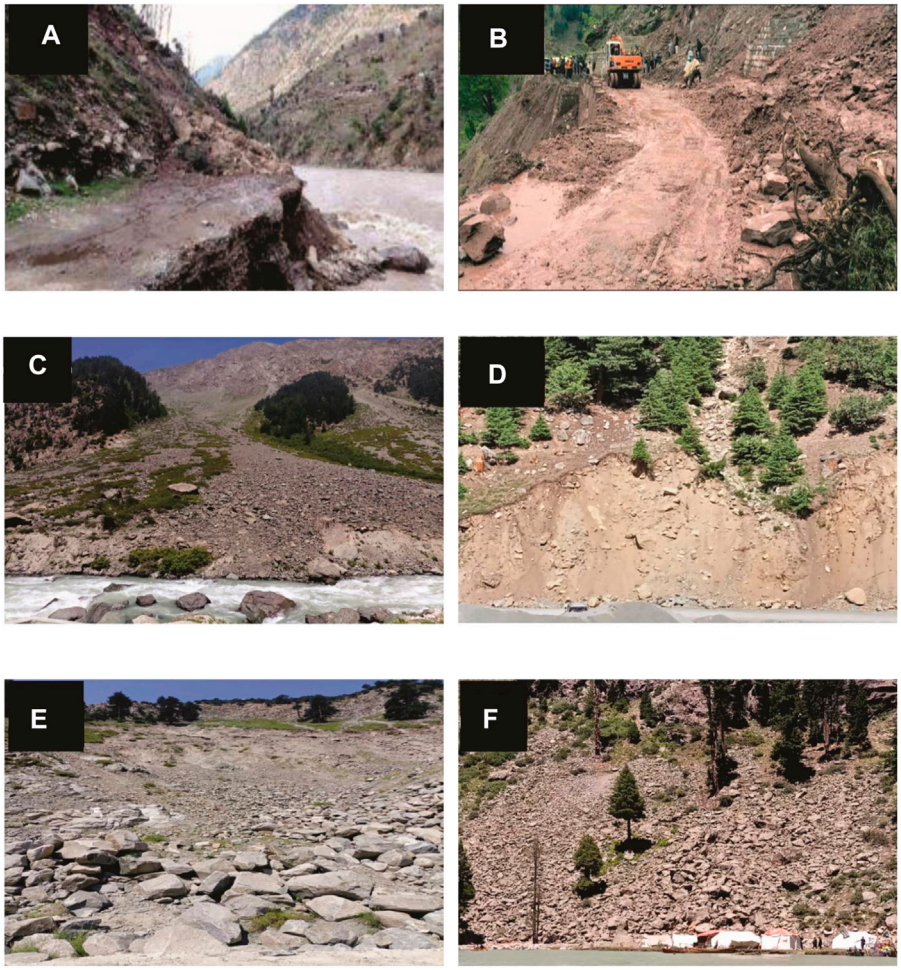


FIGURE 3
Field photographs of the study area showing the impacts of various types of landslide: **(A)** affected road and water quality; **(B)** affected road and vegetation; **(C)** affected water quality and vegetation; **(D)** affected vegetation; **(E)** affected road; **(F)** affected infrastructure.

terrain can be classified based on curvature style (Pourghasemi et al., 2013). Curvature affects surface water flow and regulates landslides (Pham et al., 2019). The curvature of the current study area was extracted from ALOS DEM and reclassified into three classes, i.e., using ArcGIS 10.8 platform as shown in Figure 4D.

3.2.5 Rainfall

Precipitation is a triggering parameter for landslide occurrence globally (Hong et al., 2017; Tariq et al., 2021a; Farhan et al., 2022). Precipitation infiltrates the pore spaces and fractures of strata and affects the landslide frequency (Dou et al., 2015). The rainfall map of the current research was generated from CHIRPS satellite data and validated with ground data of PMD, as shown in Figure 5A. The rainfall map was rescaled to a 12.5 m resolution using the ArcGIS platform.

3.2.6 Land use land cover

Many scientists have evaluated the impacts of LULC change in landslide investigation (Miller et al., 2009). LULC was computed from Landsat eight images using GEE, as shown in Figure 5B. LULC parameter was rescaled to 12.5 m spatial resolution to run model smoothly.

3.2.7 Distance from fault

The geological fault is generally considered the external causative factor in regulating the instability of beds and landslide occurrence. The tectonic map of the study area was scanned and digitized from the geological map of northern Pakistan (Searle et al., 1999), as shown in Figure 5C. The fault map was rescaled to same spatial resolution as ALOS DEM.

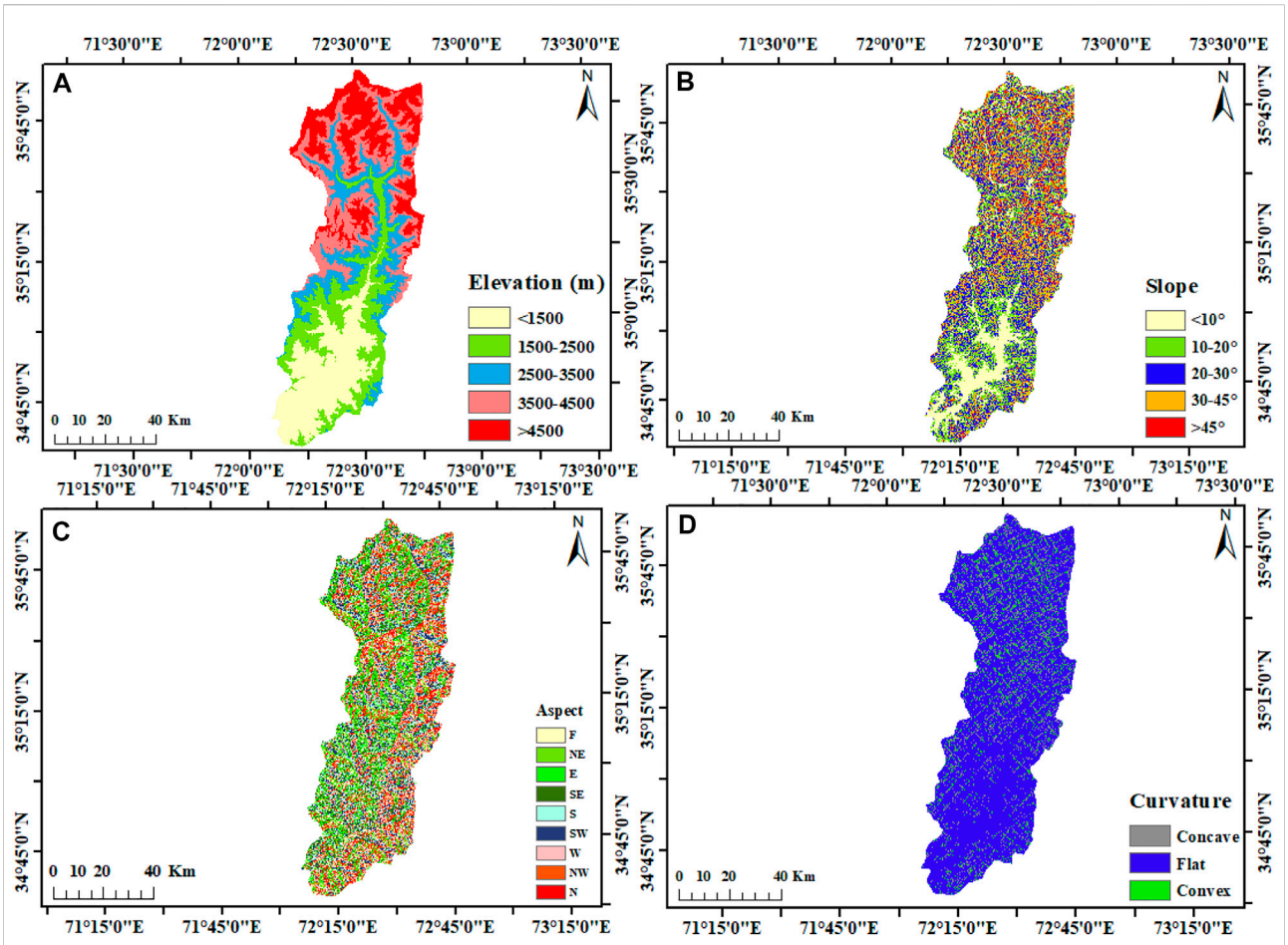


FIGURE 4
Topographic parameters derived from DEM for LSM: **(A)** elevation map; **(B)** slope map; **(C)** aspect map; **(D)** curvature map for the study area.

3.2.8 Lithology

Lithology is a crucial predisposing factor for landslides through stresses in the geological mass body (De Vallejo and Ferrer, 2011). Geology is the significant and prominent internal contributory factor in landslide susceptibility mapping (Abdollahi et al., 2019). The consolidated and compacted rocks show more resistance to mass wasting than loose rocks. A lithological map of the study area polygon vector maps was prepared and digitized from the geological map of northern Pakistan and rescaled to 12.5 m resolution, as shown in Figure 5D.

3.2.9 Distance from the road

Road construction is considered a human-induced factor for slope instability (Wu and Chen, 2009). The road network map is a polyline vector produced from the data of PKHA, as shown in Figure 5E. According to the literature, the spatial resolution of road map was rescaled to the resolution of ALOS-DEM to perform well.

3.2.10 Distance from drainage

Hydrology is a significant and influencing external parameter in the instability of strata to trigger landslides in the area. Another critical factor in the formation of landslides (Pham et al., 2015). The surface drainage network is considered one of the most active and crucial factors in landslide occurrence (Tosic et al., 2014). This parameter was computed from ALOS DEM 12.5 m resolution and divided the distance from drainage to landslide into five classes, as shown in Figure 5F.

3.3 LSM techniques

Appropriate terrain mapping is important in generating the LSM of the study area (Aslam et al., 2022b). We used WoE, FR, and IV techniques to compute LSM for the study area. The details of the mentioned methods are as follows.

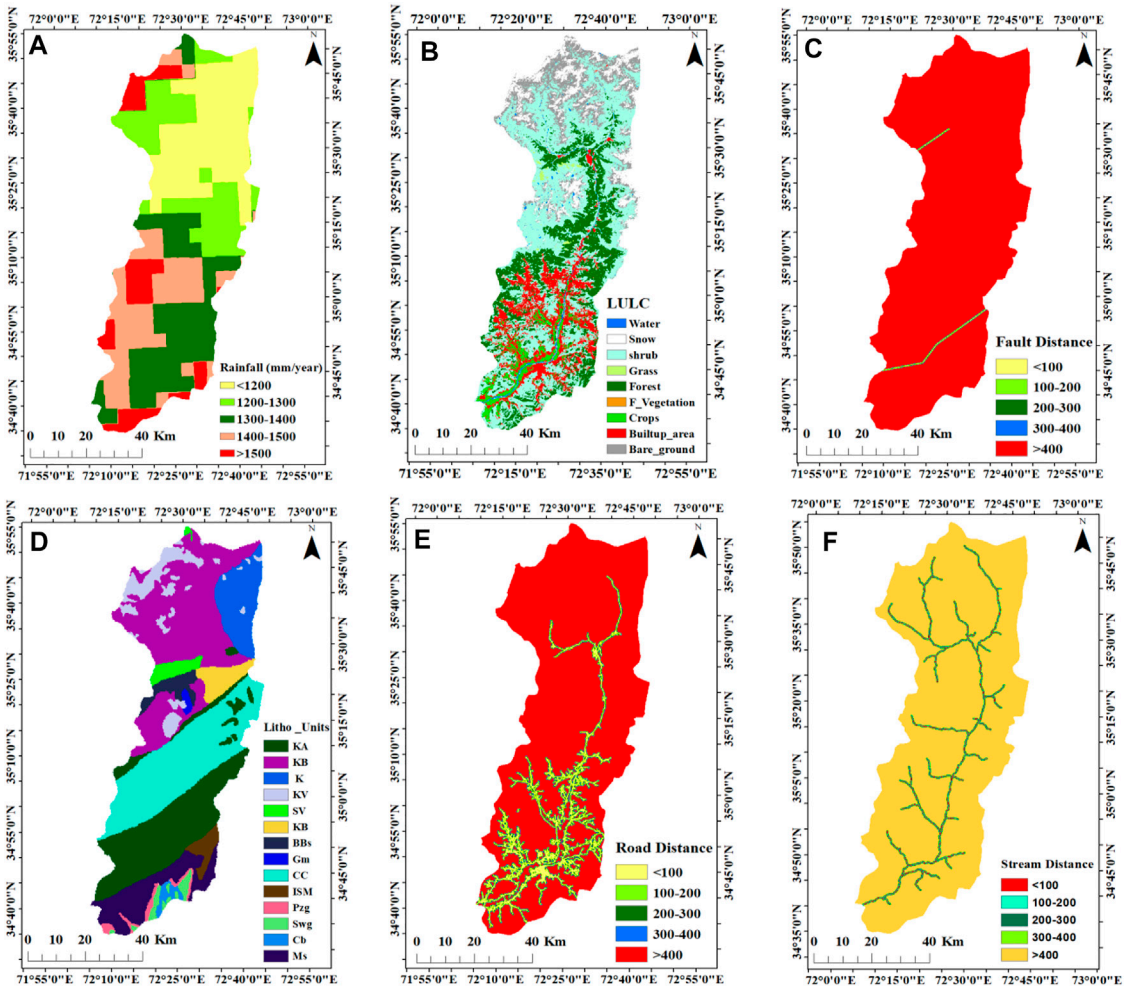


FIGURE 5
Different parameters for landslide susceptibility mapping: **(A)** precipitation map; **(B)** LULC map; **(C)** fault buffer; **(D)** lithological map; **(E)** road buffer; **(F)** stream buffer.

3.3.1 Weight of evidence

WOE is a bivariate statistical model that uses the Bayesian rule to estimate probability, adopting the concepts of prior and posterior probability (Elmoulat et al., 2015). Researchers have used this method for mineral exploration mapping. Later on, this method was widely used in landslide prediction mapping due to its authentic results compared to field and other deterministic approaches (Cao et al., 2021).

$$W^{+} = \ln \frac{h\left(\frac{\bar{B}}{D}\right)}{h\left(\frac{\bar{B}}{D}\right)} \quad (1)$$

$$W^{-} = \ln \frac{h\left(\frac{\bar{B}}{\ddot{D}}\right)}{h\left(\frac{\bar{B}}{\ddot{D}}\right)} \quad (2)$$

In this technique, positive (W^{+}) and negative (W^{-}) weights are given to different classes of causative factors and computed from Eq. 2.

In Eq. 2, h shows probability, and ln denotes the natural log. B denotes the presence of the landslide evidence parameter $\bar{B}$ on appearance of the landslide evidence parameter. Similarly, $\ddot{D}$ refers to the presence of a landslide, while $\ddot{D}$ is the absence of a landslide.

We further used Eq. 3 to evaluate the impacts of causative factors on LS occurrence.

$$W^{+} = \ln \frac{NP_i x_1}{NP_i x_2} + \frac{NP_i x_2}{NP_i x_3} + \frac{NP_i x_3}{NP_i x_4} \quad (3)$$

$$W^{-} = \ln \frac{NP_i x_3}{NP_i x_1} + \frac{NP_i x_2}{NP_i x_4} + \frac{NP_i x_3}{NP_i x_4} \quad (4)$$

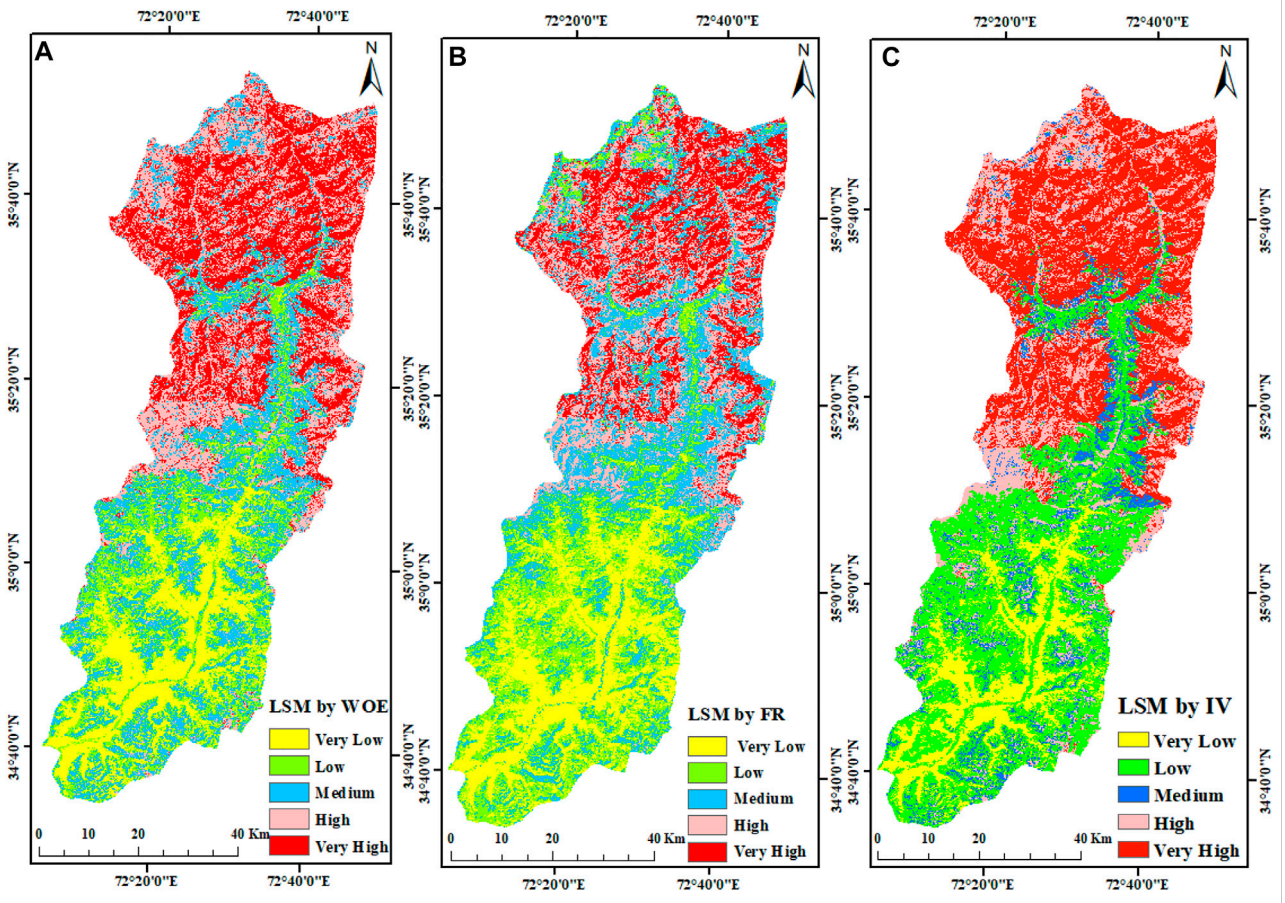


FIGURE 6
Landslide susceptibility mapping derived from bivariate models: **(A)** LSM by WOE; **(B)** LSM by FR; **(C)** LSM by IV.

where NP_ix_1 number of pixels shows the existence of predisposing factors and landslides, $NP_ix_2$is the absence of landslides predisposing parameter and presence of landslide, NP_ix_3 is the presence of the contributing factors to landslides and the non-existence of landslides, and $NP_ix_4$is the absence of both landslides and landslides predisposing factors. The final LSM of WOE is shown in Figure 6A.

3.3.2 Frequency ratio

The FR model is considered to be among the best bivariate statistical models for use in computing the spatial association between two variables (Oh et al., 2017). This statistical method is a reliable experimental technique to produce LSM in the research area (Fayez et al., 2018). Eq. 5 is used to calculate the FR for each factor.

$$FR = \frac{N_iP_x/N}{N_il_Q/Nl} \quad (5)$$

where *FR* = frequency ratio, N_iP_x = number of pixels in each landslides conditioning factor class, N = number of all pixels in the study area, N_il_P = number of landslide pixels in each landslide conditioning factor, and *Nl* = number of all landslide pixels in the study area.

The following mathematical representation is used to generate LSI for the region of interest.

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

The expression FR_{ij} is frequency ratio value for the "j" class of factor "i", and n is the total number of factors. After performing these steps, the LSM map was produced, as shown in Figure 6B.

3.3.3 Information value

In this work, the IV method generates the LSM of the study area. This statistically based GIS technique predicts the spatial association between landslide inventory and classes of predisposing factors (Li et al., 2021).

This analysis can be achieved through the following calculation:

TABLE 2 Detailed analysis of different causative parameters with landslide events using bivariate models (IV, FR, and WoE) to understand the impacts of each class in LSM.

Parameters	Class	No. of pixels in a class	No. of landslide pixels in a class	W^+	W^-	WC	% of pixels in a class	% of LS pixels in a class	(FR)	IV = log (A/B)
Elevation	<1,500	7,738,778	756	−4.78	0.26	−5.04	22.52	0.190	0.01	−4.77
	1,500–2,500	6,531,545	4,635	−2.80	0.20	−3.00	19.01	1.168	0.06	−2.79
	2,500–3,500	5,878,979	101,163	0.42	−0.11	0.53	17.11	25.486	1.49	0.42
	3,500–4,500	7,471,119	150,000	0.82	−0.23	0.79	21.74	37.790	1.74	0.55
	>4,500	6,738,623	140,374	0.21	−0.22	0.82	19.61	35.365	1.80	0.59
Slope	<10°	5,900,546	16,596	−1.42	−0.14	−1.57	17.17	4.18	0.24	−1.41
	10–20	8,400,816	69,688	−0.33	−0.08	−0.42	24.45	17.56	0.72	−0.33
	20–30	10,285,258	124,369	0.40	−0.04	0.26	29.93	31.33	1.05	0.17
	30–45	7,243,458	141,275	0.175	0.14	0.54	21.08	35.59	1.69	0.43
	>45	2,528,966	45,000	0.43	0.08	0.48	7.36	11.34	1.54	0.40
Aspect	F	3,547,098	15,766	−0.96	0.07	−1.03	10.32	3.97	0.38	−0.96
	NE	3,747,398	53,686	0.22	−0.03	0.25	10.91	13.53	1.24	0.22
	E	3,819,815	69,991	0.47	−0.08	0.55	11.12	17.63	1.59	0.46
	SE	4,111,079	77,868	0.50	−0.09	0.59	11.97	19.62	1.64	0.49
	S	4,011,237	62,890	0.31	−0.05	0.36	11.67	15.84	1.36	0.31
	SW	4,214,290	61,685	0.24	−0.04	0.28	12.27	15.54	1.27	0.24
	W	3,938,948	39,437	−0.14	0.02	−0.16	11.46	9.94	0.87	−0.14
	NW	3,613,212	11,340	−1.31	0.08	−1.39	10.52	2.86	0.27	−1.30
	N	3,355,967	4,265	−2.22	0.09	−2.31	9.77	1.07	0.11	−2.21
Curvature	Concave	3,828,011	287,611	1.94	−1.18	3.12	11.14	72.46	6.50	1.87
	Flat	28,186,139	66,509	−1.60	1.58	-3.17	82.03	16.76	0.20	−1.59
	Convex	2,344,894	42,808	0.46	−0.04	0.51	6.82	10.78	1.58	0.46
Precipitation	<1,200	9,463,996	197,974	0.60	−0.37	0.98	27.54	49.88	1.81	0.59
	1,200–1,300	6,471,282	120,882	0.49	−0.16	0.64	18.83	30.46	1.62	0.48
	1,300–1,400	8,295,914	49,103	−0.67	0.15	−0.82	24.14	12.37	0.51	−0.67
	1,400–1,500	6,692,712	16,976	−1.52	0.18	−1.70	19.48	4.28	0.22	−1.52
	>1,500	3,441,168	11,957	−1.40	0.08	−1.29	10.01	3.01	0.30	−1.20
LULC	Water	260,922	448	−1.92	0.01	−1.92	0.76	0.11	0.15	−1.91
	Forest	6,891,142	387	−5.34	0.23	−5.56	20.07	0.10	0.00	−5.33
	Grass	213,081	174	−2.66	0.01	−2.67	0.62	0.04	0.07	−2.65
	Flood vegetation	45	1	0.66	0.00	0.66	0.00	0.00	1.92	0.65
	Crops	1,257,688	154	−4.56	0.04	−4.60	3.66	0.04	0.01	−4.55
	shrub	14,895,916	234,000	0.31	−0.32	0.64	43.39	58.95	1.36	0.31
	Builtup_area	3,675,808	1,139	−3.63	0.11	−3.74	10.71	0.29	0.03	−3.62
	Bare ground	4,185,037	147,827	1.14	−0.34	1.48	12.19	37.24	3.05	1.12
	Snow	2,947,648	12,798	−0.99	0.06	−1.04	8.59	3.22	0.38	−0.98
Distance to fault	<100	75,904	10	−4.49	0.00	−4.49	0.22	0.00	0.01	−4.474
	100–200	76,090	73	−2.50	0.00	−2.50	0.22	0.02	0.08	−2.489
	200–300	76,257	91	−2.28	0.00	−2.28	0.22	0.02	0.10	−2.270
	300–400	76,440	215	−1.42	0.00	−1.42	0.22	0.05	0.24	−1.413
	>400	34,061,693	396,732	0.01	−2.21	2.22	99.11	99.90	1.01	0.008
Lithology	Ka	6,347,601	3,158	−3.17	0.20	−3.37	34.41	0.80	0.0231	−3.16
	KB	9,948,821	235,797	0.72	−0.56	1.28	53.93	59.41	1.1015	0.71
	EC	1,185,411	86,665	1.90	−0.21	2.11	6.43	21.84	3.3978	1.83

(Continued on following page)

TABLE 2 (*Continued*) Detailed analysis of different causative parameters with landslide events using bivariate models (IV, FR, and WoE) to understand the impacts of each class in LSM.

Parameters	Class	No. of pixels in a class	No. of landslide pixels in a class	W^+	W^-	WC	% of pixels in a class	% of LS pixels in a class	(FR)	IV = log (A/B)
	GI	1,439,183	39,579	0.87	−0.06	0.93	7.80	9.97	1.2781	0.86
	UV	1,180,316	12,271	−0.12	0.00	−0.12	6.40	3.09	0.4832	−0.12
	KV	957,401	2,726	−1.42	0.02	−1.44	5.19	0.69	0.1323	−1.41
	BBS	916,768	14,635	0.32	−0.01	0.33	4.97	3.69	0.7419	0.31
	GM	155,417	4	−6.13	0.00	−6.13	0.84	0.00	0.0012	−6.12
	CC	6,876,463	2015	−3.70	0.22	−3.92	37.28	0.51	0.0136	−3.69
	ISM	352,208	4	−6.95	0.01	−6.96	1.91	0.00	0.0005	−6.94
	PZG	757,560	6	-7.31	0.02	-7.33	4.11	0.00	0.0004	−7.30
	SWG	800,352	11	−6.76	0.02	−6.78	4.34	0.00	0.0006	−6.75
	CB	268,165	13	−5.50	0.01	−5.50	1.45	0.00	0.0023	−5.49
	Q	463,000	3	−7.51	0.01	−7.52	2.51	0.00	0.0003	−7.50
	MS	2,313,318	5	−8.61	0.07	−8.68	12.54	0.00	0.0001	−8.60
Distance to road	<100	2,078,393	4,753	−.63	0.05	−1.68	6.05	1.20	0.20	−1.62
	100–200	1,476,459	4,014	−1.46	0.03	−1.49	4.30	1.01	0.24	−1.45
	200–300	1,163,424	4,415	−1.12	0.02	−1.14	3.39	1.11	0.33	−1.11
	300–400	998,097	4,137	−1.03	0.02	−1.05	2.90	1.04	0.36	−1.02
	>400	28,650,011	379,609	0.14	−1.35	1.49	83.37	95.64	1.15	0.14
Distance to stream	<100	810,632	25,395	0.94	−0.04	0.98	2.36	6.40	2.71	0.93
	100–200	1,017,643	23,000	0.61	−0.03	0.63	2.96	5.79	1.96	0.60
	200–300	1,222,348	21,800	0.37	−0.02	0.38	3.56	5.49	1.54	0.36
	300–400	1,393,655	17,000	−0.02	0.00	−0.02	4.06	4.28	1.06	−0.02
	>400	29,922,106	339,530	−0.09	0.46	−0.56	87.07	85.54	0.98	−0.09

$$W = \log \frac{\frac{MQox(R_o)}{MQ_ox(R_o)}}{\sum \frac{MQ_ox(R_o)}{\sum MQ_ox(R_o)}} \quad (7)$$

where W denotes the weight of the causative factor for landslides.

$Mox(R_o)$ shows the number landslide of pixels within class "o"; $MQ_ox(M_o)$ is the number of all pixels within class "o"; $MQ_Ox(R_O)$ is the total number of landslides pixels; and $\sum MQ_Ox(M_i))$ is used for the total number of pixels in the study area. The LSI can be generated for the study area using the following formula:

$$LSI = W_E + W_S + W_A + W_C + W_{LULC} + W_L + W_P + W_F + W_R + W_D \quad (8)$$

where W_E = weight of elevation, W_S = weight of slope, W_A = weight of aspect, W_C weight of curvature, W_{LULC} = weight of landuse landcover, W_L = weight of lithology, W_F = weight of fault, W_R = weight of road, W_P = weight of rainfall, and W_D = weight of stream network. The LSM of IV is given in Figure 6C.

4 Results

In this research, 495 landslide events were detected by the integrated interpretation of satellite imageries and ground-based data in the study area. The spatial distribution of landslide events is given in Figure 1. These landslide events consist of different types of landslides, i.e., mudflow, debris, rockfall, rockslide, topple, and creep. In the current study, three bivariate models are applied to produce the LSM of the study area. The details of the results for each model are shown in Table 2. The comprehensive, detailed description of Table 2 is as follows.

The elevation parameters used in the current study represent a strong association with landslide events. The most influential class of elevation is >4,500 m, followed by 3,500–4,500 m and 2,500–3,500 m. The < 1,500 m class of elevation is less susceptible. The independent variable, i.e., the slope, is considered the crucial factor in the present study. The slope factor is influential up to 45° because as slope increases, landslide occurrence also increases, but above 45°, landslide activity declines with increasing slope, as shown in Table 2. The

results revealed that 30°–45° is the most susceptible class of slope while, slopes of <10° are the most resistant to landslide, followed by 10°–20° class of slope. Table 2 indicates that SE is the most crucial class of aspects, followed by E, S, and SW. The tabulated results explained that concave structure is the critical class of landslides, as noted in Table 2. In the present study, a rainfall map was generated from CHIRPS data, validated with ground-based data, and then reclassified into five classes to evaluate the association of rainfall parameters with landslide events. As shown in Table 2 for precipitation, the results indicated that rainfall is a significant factor for landslides. The results revealed that <1,200 mm/year precipitation class is the censorious class for landslide, followed by 1,200–1,300, 1,300–1,400, 1,400–1,500, and >1,500 mm/year. In the current research, the tabulated results clearly explained that every class of LULC has diverse impacts on landslide events. The results of Table 2 show that the barren land and flooded vegetation of the present study area is most susceptible to landslides. The forest and built-up area show greater resistance to landslides in the study area than other classes of LULC.

The results of the present study demonstrate that faults have no direct influence on landslide occurrence, as shown in Table 2. The results illustrate that very a smaller number of landslide pixels occurred in a <100 m buffer zone near the fault, i.e., values of −4.49, 0.01, and −4.47 for WOE, FR, and IV, respectively. The most vulnerable buffer zone of a geological fault is >400 m, as most landslide events occur in this buffer region. The WOE, FR, and IV models for the >400 m buffer are 2.22, 1.01, and 0.008, respectively.

The results prove that lithology is a significant causative factor for landslide investigation. As shown in Table 2, EC is the most susceptible geological formation for landslides, followed by GI and KB. The output results of bivariate models for road association with landslide events, as shown in Table 2, indicate that the road network has no direct impact on the instability of strata. The <100 m buffer zone of road network is not susceptible to landslides in the study area. No direct association can be observed between the dependent variable and the road network. The >400 m buffer zone of the road network is most susceptible to landslides with values of 1.49, 1.15, and 0.14 for the WOE, FR, and IV models.

The results for landslide events and stream networks show that both variables feature a substantial direct association with each other, because in this study, we observed that most landslide events and landslide pixels occurred near streams, while smaller numbers of landslide events were observed far from drainage. The bivariate statistical results revealed that a <100 m buffer from the stream network is the most susceptible class to landslide occurrence followed by 100–200 m and 200–300 m. The WOE, FR, and IV models values of −0.56, 0.98, and −0.09 revealed that both dependent variable and independent variables in the >400 m buffer zone are not susceptible to landslide.

The tabulated analysis concluded that the concave structure is the most crucial class for landslide occurrence, followed by lithology, barren land, stream network, and elevation. The influential factors for the WOE results mentioned above are 3.12, 2.11, 1.48, 0.98, and 0.82, respectively. The FR model results for concave, lithology, bare ground, stream network, and elevation are 6.50, 3.39, 3.05, 2.71, and 1.80, respectively. The IV model values for concave, lithology, bare ground, stream network, and elevation are 1.87, 1.83, 1.12, 0.93, and 0.59.

This study used GIS-based statistical analysis of various causative factors with landslide events in ArcGIS 10.8 to produce the LSM of the research area, as shown in Figure 6. These were reclassified into five classes, i.e., very low, low, high, and very high. The LSM produced by WOE, as shown in Figure 6A, indicated that the major portion of the north region of the study area is highly susceptible to landslides, followed by high and moderate susceptibility class of landslides, while the south region of the LSM is occupied has low and low susceptibility classes. The south region of the study area is safe according to results of WOE model. The final LSM generated by FR model, as shown in Figure 6B, reveals that the most northern area of LSM is exposed to very high, high, moderate susceptibility. By contrast to the WOE model in the results for this technique, we observe some low and very low susceptibility zones in the region of the study area. The resulting map of the IV model is different from both mentioned models. In this LSM map, the north region is mostly susceptible to the very high and high class. The very low zone is not mentioned in the north region, and the moderate class is negligible in this LSM relative to WOE and FR. The WOE and FR on this LSM of the IV model contains low and very low class of susceptibility in the north region, which was not identified in the previous two models. The south portion of LSM, generated by the IV model, consists of very low and low susceptibility class and contains some high zones that are not mentioned in the previous model amps. Beyond the north and south zones of the LSM, the central region of the WOE and FR models are mostly susceptible to high and moderate susceptibilities, while some areas have low and very high classes as well. In the LSM generated by IV model the central region is mostly susceptible to low class. According to the IV model, the central region is safe.

To validate the performance of bivariate models, we used the AUROC method. The validation graphs of WOE, FR, and IV are shown in Figure 7 using 30% of landslide inventory data. AUC chart for the WOE model showed that the SRC and PRC values of the model are 67.3% and 87%, respectively. The SRC and PRC graphs for the WOE model are shown in Figures 7A,B. The validation charts of the FR bivariate model illustrate values of 0.93 and 0.95 for the SRC and PRC, respectively. Based on training and validation data, these SRC and PRC

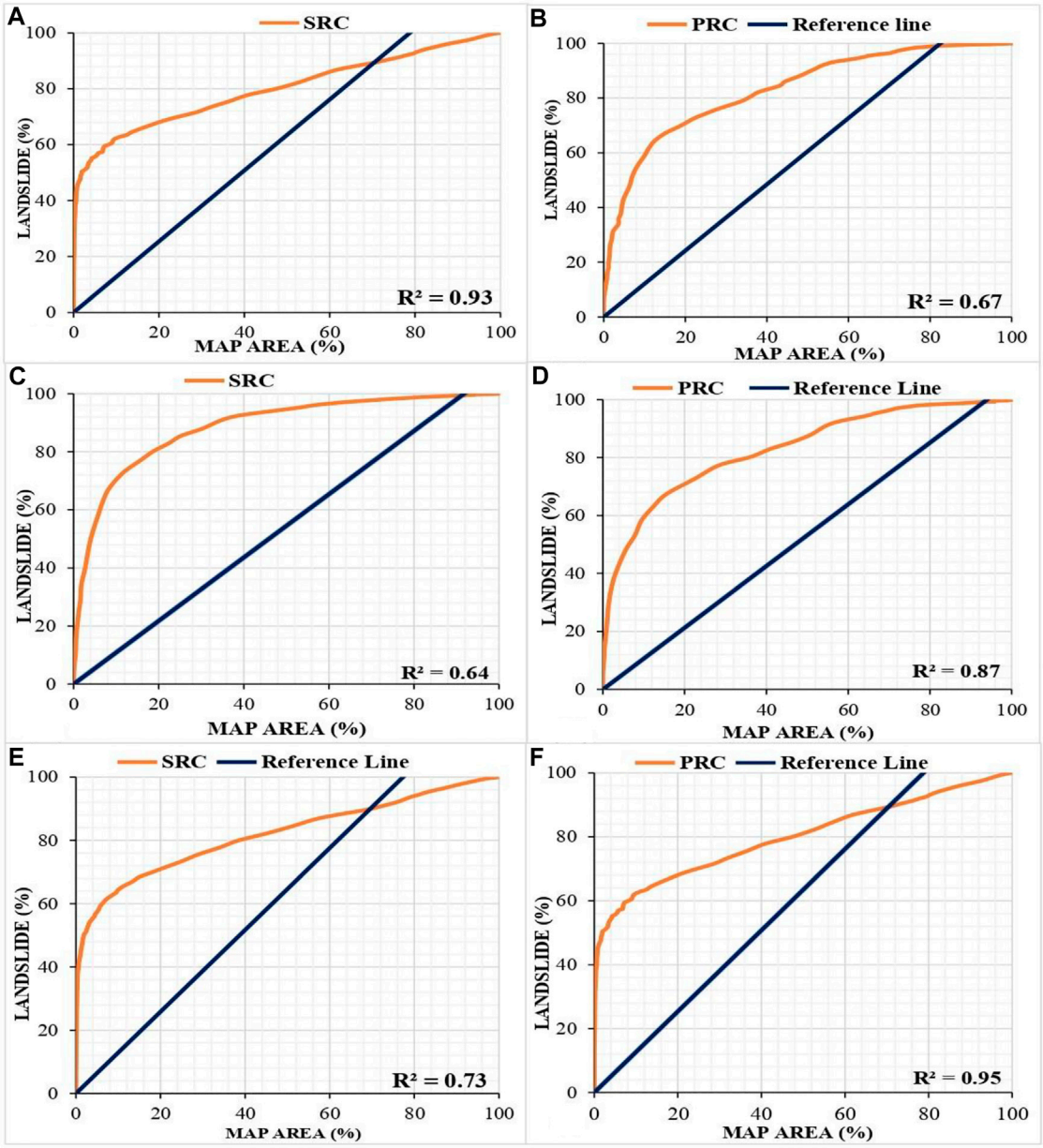


FIGURE 7
Success rate curve (SRC) and predicted rate curve (PRC) for WOE, FR, and IV model: **(A)** SRC for WOE; **(B)** PRC for WOE; **(C)** SRC for FR model; **(D)** PRC for FR model; **(E)** SRC for IV; **(F)** PRC for IV.

values can be represented in percentage format, i.e., 93% and 95%, respectively. The SRC and PRC graphs are illustrated in Figures 7C,D. The visual record for the IV model revealed that the chart values were 0.64 and 0.73 for SRC and PRC, respectively meant that the model accuracy is 64% and 73.85%. The SRC and PRC graphical representation for the IV model is shown in Figures 7E,F. The validation outcomes of WOE, FR, and IV indicate that the FR is a reliable model to produce LSM for the present study area.

5 Discussion

Landslide development is a multifaceted phenomenon because it is triggered and predisposed by various natural and human-induced factors (Chen et al., 2019). In this study, LSM was generated using geospatial techniques based on landslide events and predisposing factors (elevation, slope, aspect, curvature, precipitation, LULC, distance to fault, lithology, distance to road, and distance to streams) to mitigate current

and future hazard impacts. All 10 causative factors were considered important because they were selected and prepared according to the literature, but some parameters are more influential than others. First of all, we prepared an inventory map using satellite and ground-based data. The topographic factors of altitude, slope, aspect, and curvature are reflected as the influential parameters for landslides (Alkhasawneh et al., 2013; Tariq et al., 2020; Ghaderizadeh et al., 2021). This paper focused on topographic causative factors that are associated with landslide events, and a tabulated explanation is given in Table 2. The outcome results of topographic parameters in this study prove that these are the most influential independent variables for landslide occurrence. The results section of this paper indicates that curvature, lithology, barren ground, stream network, and elevation are the most significant parameters, while the contribution of other factors also can never be ignored. The concave curvature is more strongly associated with landslide events than convex structures (Xu et al., 2014). The results of the present study clearly showed that concave structure is mainly affected by landslides than convex curvature.

The precipitation is the most influential external triggering factor for landslides but also depend on the lithology, slope, and LULC of the study area (Silalahi et al., 2019). In the present study heavy precipitation occurred in the hilly region which have compacted lithology and vegetation cover, so the area was not affected by landslide hazard. In the same study area some region have low precipitation with major landslide because low precipitation area was occupied by loose lithology and barren land.

The LSM literature indicates that lithology is often considered an influential factor (Segoni et al., 2020). The findings this analysis agree that lithology is an influential and significant internal factors in LSM. The literature suggests that barren class of land cover is the most influential triggering factor in landslide occurrence (Khan et al., 2019). The findings of this study confirm that barren land is the susceptible class of land cover for a landslide. The last causative factor found in this research but not the least is drainage network, which is an external influential causative factor in landslide occurrence (Pradhan et al., 2012). This study revealed strongly positive correlation of drainage with landslides. During a field survey, we validated ground truth information with GIS-based models and concluded that all landslide-affecting factors contribute to landslide occurrence, but curvature, lithology, barren ground, stream network, and elevation are the most trigging causative factors of landslides in the current study area.

In the current paper, lithology was proven to have a crucial role in landslide occurrence. The literature indicates that bare ground is more susceptible to landslide occurrence (Khan et al., 2019). This paper also showed that barren land is the most susceptible class to landslides, with values of 1.48, 3.05, and 1.12 for WOE, FR, and IV. The drainage network is an influential factor in a landslide (Pradhan et al., 2012), and in this study, drainage is strongly associated with landslide events.

The outcomes demonstrated that the findings of success rate curve (SRC) of the WOE, FR, and IV models were 67%, 93%, and 64%, respectively, while the prediction rate curve (PRC) of the three models were 87%, 95%, and 73%, respectively. In the current research, the FR model has 93% and 95% values for SRC and PRC, respectively, and produces the best prediction mapping of the landslide, as shown in Figure 6B. The WOE model also made a good prediction map for the study area. In contrast, the IV model has not produced satisfactory results.

Geoscientists have adopted various statistical models in numerous regions and attained different findings. This type of discrepancy is mostly due to weight differences, along with the selection of models and causative factors. Therefore, landslide detection, the selection of landslide-affecting factors, and appropriate models for the study area is very important for reducing uncertainty in model processing and prediction (Pham et al., 2019). This is because GIS-based models depend upon the reliability and quality of input data. In this research, we used quantitative models and achieved superior accuracy to qualitative and semi-quantitative methods.

6 Conclusion

This research was designed to generate an LSM of the study area using geospatial techniques to mitigate the consequences of hazards. Three GIS-based statistical models, i.e., WOE, FR, and IV, were applied in the current research to compute the association of dependent variables (landslide causative factors) and dependent variables (landslide events/inventory). The study results delivered significant evidence concerning landslide existence in the study area. The results explained landslides instigated by various causative factors in the study area. This study was executed to identify regions susceptible to landslides and classify them into deficient, low, medium, high, and very high zones to alleviate their consequences, using geospatial techniques. This research was conducted to evaluate the association of causative factors with landslide occurrence. These parameters were topographic, geologic, hydrologic, climatic, and geomorphic. The three GIS-based statistical models were used to investigate the association of landslide occurrence with causative factors to produce LSM.

From the above discussion and due to the results of analysis, we concluded that the highest value for LSM in the current research area by bivariate analysis was the value of curvature, lithology, barren ground, stream network, and elevation. The most susceptible class of curvature was a concave structure having values of 3.12, 6.50, and 3.12 values for the WOE, FR and IV models, respectively. The WOE, FR, and IV models for flat class were −3.17, 0.20, and −1.59, respectively, which shows that it is the most less susceptible class. The results revealed that the EC is the most susceptible formation in the current study,

having association values of 2.11, 3.39, and 1.83 for the WOE, FR, and IV models, respectively. The results concluded that barren land the association values were WOE, FR, and IV are 1.48, 3.05, and 1.12. The tabulated analysis revealed that <100 m buffer zone of the stream network was most susceptible to landslide occurrence. The analytical results for this buffer zone of WOE, FR, and IV were 0.98, 2.71, and 0.93, respectively.

The validation results, i.e., SRC and PRC for the WOE, FR, and IV models, were 0.67, 0.87, 0.93, 0.95, 0.64, and 0.73, respectively. The validation results revealed that FR model is a credible method for the LSM, as this model showed accuracies of 0.93 and 0.95 for SRC and PRC, respectively. It can be concluded that GIS-based statistical modeling is the most reliable, flexible, and authentic method for generating LSM. Various organizations can use the final LSM of the bivariate models to reduce the effects of landslide hazards in the study area.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

Fl: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. SR: Writing – review and editing, Visualization. BG: Writing – review and editing, Visualization. AT: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Supervision, Project administration. SUS: Writing – review and editing, Visualization. MN: Writing – review and editing, Visualization. MLH: Writing – review and editing, Visualization. NUA: Writing – review and editing, Visualization. QL: Writing – review and editing, Visualization, Project administration, Funding. LL: Writing – review and editing, Visualization, Project administration, Funding. MS: Writing – review and editing, Visualization. MA: Writing – review and editing, Visualization.

Funding

This work is supported by the National Natural Science Foundation of China (grant no. 41901292 and grant no. 42071321).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Pham, B. T., Bui, D. T., Prakash, I., and Dholakia, M. B. (2015). Landslide susceptibility assessment at a part of uttarakhand Himalaya, India using GIS – based statistical approach of frequency ratio method. *Int. J. Eng. Res.* 4. doi:10.17577/ijertv4is110285

Abbas, I., Liu, J., Amin, M., Tariq, A., and Tunio, M. H. (2021). Strawberry fungal leaf scorch disease identification in real-time strawberry field using deep learning architectures. *Plants* 10, 2643. doi:10.3390/plants10122643

Abdollahi, S., Pourghasemi, H. R., Ghanbarian, G. A., and Safaeian, R. (2019). Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions. *Bull. Eng. Geol. Environ.* 78, 4017–4034. doi:10.1007/s10064-018-1403-6

Alkhasawneh, M. S., Ngah, U. K., Tay, L. T., Mat Isa, N. A., and Al-Batah, M. S. (2013). Determination of important topographic factors for landslide mapping analysis using MLP network. *Sci. World J.* 2013, 1–12. doi:10.1155/2013/415023

Aslam, B., Maqsoom, A., Khalil, U., Ghorbanzadeh, O., Blaschke, T., Farooq, D., et al. (2022a). Evaluation of different landslide susceptibility models for a local scale in the chitral district, northern Pakistan. *Sensors* 22, 3107. doi:10.3390/s22093107

Aslam, B., Zafar, A., and Khalil, U. (2022b). *Comparative analysis of multiple conventional neural networks for landslide susceptibility mapping*. Netherlands: Springer. doi:10.1007/s11069-022-05570-x

Atta-ur-Rahman, A., and Khan, A. N. (2011). Analysis of flood causes and associated socio-economic damages in the Hindukush region. *Nat. Hazards* 59, 1239–1260. doi:10.1007/s11069-011-9830-8

Bahadar, I., Shafique, M., Khan, T., Tabassum, I., and Ali, M. Z. (2015). Flood hazard assessment using hydro-dynamic model and GIS/RS tools: A case study of babuzai-kabal tehsil swat basin, Pakistan. *J. Himal. Earth Sci.* 48, 129–138.

Baloch, M. Y. J., Zhang, W., Chai, J., Li, S., Alqurashi, M., Rehman, G., et al. (2021). Shallow groundwater quality assessment and its suitability analysis for drinking and irrigation purposes. *WaterSwitzerl.* 13, 3361–3425. doi:10.3390/w13233361

Baqa, M. F., Chen, F., Lu, L., Qureshi, S., Tariq, A., Wang, S., et al. (2021). Monitoring and modeling the patterns and trends of urban growth using urban sprawl matrix and CA-markov model: A case study of karachi, Pakistan. *Land* 10 (7), 700. doi:10.3390/land10070700

Barredo, J. I., Benavides, A., Hervás, J., and Van Westen, C. J. (2000). Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. *Int. J. Appl. Earth Obs. Geoinf.* 2000, 9–23. doi:10.1016/s0303-2434(00)85022-9

Cao, Y., Wei, X., Fan, W., Nan, Y., Xiong, W., and Zhang, S. (2021). Landslide susceptibility assessment using the weight of evidence method: A case study in xunyang area, China. *PLoS One* 16, 02456688–e245718. doi:10.1371/journal.pone.0245668

Chalkias, C., Ferentinou, M., and Polykretis, C. (2014). GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece. *Geosci. (Basel).* 4, 176–190. doi:10.3390/geosciences4030176

Chen, J., Du, L., and Guo, Y. (2021). Label constrained convolutional factor analysis for classification with limited training samples. *Inf. Sci. (N. Y).* 544, 372–394. doi:10.1016/j.ins.2020.08.048

Chen, W., Panahi, M., Tsangaratos, P., Shahabi, H., Ilia, I., Panahi, S., et al. (2019). Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *Catena* 172, 212–231. doi:10.1016/j.catena.2018.08.025

Chen, Z., Liu, Z., Yin, L., and Zheng, W. (2022). Statistical analysis of regional air temperature characteristics before and after dam construction. *Urban Clim.* 41, 101085. doi:10.1016/j.uclim.2022.101085

Chimidi, G., Raghuvanshi, T. K., and Suryabhagavan, K. V. (2017). Landslide hazard evaluation and zonation in and around gimbi town, Western Ethiopia—A GIS-based statistical approach. *Appl. Geomat.* 9, 219–236. doi:10.1007/s12518-017-0195-x

Choi, J., Oh, H. J., Lee, H. J., Lee, C., and Lee, S. (2012). Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Eng. Geol.* 124, 12–23. doi:10.1016/j.enggeo.2011.09.011

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., and Chacón, J. (2012). Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: Application to the river Beiro basin (Spain). *Nat. Hazards Earth Syst. Sci.* 12, 327–340. doi:10.5194/nhess-12-327-2012

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., and Paudyal, P. (2008). Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence. *Geomorphology* 102, 496–510. doi:10.1016/j.geomorph.2008.05.041

De Vallejo, L. G., and Ferrer, M. (2011). *Geological engineering*. London: Geological Engineering. CRC Press.

Dou, J., Bui, D. T., Yunus, A. P., Jia, K., Song, X., Revhaug, I., et al. (2015). Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. *PLoS One* 10, e0133262. doi:10.1371/journal.pone.0133262

Elmoulat, M., Ait Brahim, L., Mastere, M., and Ilham Jemmah, A. (2015). Mapping of mass movements susceptibility in the zoumi region using satellite image and GIS technology (Moroccan rif). *Int. J. Sci. Eng. Res.* 6.

Farhan, M., Moazzam, U., Rahman, G., Munawar, S., Tariq, A., Safdar, Q., et al. (2022). Trends of rainfall variability and drought monitoring using standardized precipitation index in a scarcely gauged basin of northern Pakistan. *Water* 14, 1132. doi:10.3390/w14071132

Fayez, L., Pazhman, D., Pham, B. T., Dholakia, M. B., Solanki, H. A., Khalid, M., et al. (2018). Application of frequency ratio model for the development of landslide susceptibility mapping at part of uttarakhand state, India. *Int. J. Appl. Eng. Res.* 13, 6846–6854.

Fu, C., Cheng, L., Qin, S., Tariq, A., Liu, P., Zou, K., et al. (2022). Timely plastic-mulched cropland extraction method from complex mixed surfaces in arid regions. *Remote Sens. (Basel).* 14, 4051. doi:10.3390/rs14164051

Ghaderizadeh, S., Abbasi-Moghadam, D., Sharifi, A., Zhao, N., and Tariq, A. (2021). Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 7570–7588. doi:10.1109/JSTARS.2021.3099118

Girma, F., Raghuvanshi, T. K., Ayenew, T., and Hailemariam, T. (2015). Landslide hazard zonation in Ada Berga District, Central Ethiopia–a GIS based statistical approach. *J. Geom.* 9, 25–38.

Guo, Y., Yang, Y., Kong, Z., and He, J. (2022). Development of similar materials for liquid-solid coupling and its application in water outburst and mud outburst model test of deep tunnel. *Geofluids* 2022, 1–12. doi:10.1155/2022/8784398

Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P. (1995). Stern, 1991] and one of the NAVZ (Northern Austral Vol- canic Zone) volcanic events [3010 yr. *Geomorphology* 13, 1995.

Hong, H., Chen, W., Xu, C., Youssef, A. M., Pradhan, B., and Tien Bui, D. (2017). Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int.* 32, 1–16. doi:10.1080/10106049.2015.1130086

Hu, P., Sharifi, A., Tahir, M. N., Tariq, A., Zhang, L., Mumtaz, F., et al. (2021). Evaluation of vegetation indices and phenological metrics using time-series modis data for monitoring vegetation change in Punjab, Pakistan. *WaterSwitzerl.* 13, 2550–2615. doi:10.3390/w13182550

Hussain, S., Lu, L., Mubeen, M., Nasim, W., Karuppannan, S., Fahad, S., et al. (2022). Spatiotemporal variation in land use land cover in the response to local climate change using multispectral remote sensing data. *Land* 11, 595. doi:10.3390/land11050595

Imran, M., Ahmad, S., Sattar, A., and Tariq, A. (2022). Mapping sequences and mineral deposits in poorly exposed lithologies of inaccessible regions in Azad Jammu and Kashmir using SVM with ASTER satellite data. *Arab. J. Geosci.* 15, 538. doi:10.1007/s12517-022-09806-9

Khan, H., Shafique, M., Khan, M. A., Bacha, M. A., Shah, S. U., and Calligaris, C. (2019). Landslide susceptibility assessment using Frequency Ratio, a case study of northern Pakistan. *Egypt. J. Remote Sens. Space Sci.* 22, 11–24. doi:10.1016/j.ejrs.2018.03.004

Khanchoul, K., Balla, F., and Othmani, O. (2020). Assessment of soil erosion by rusle model using gis: A case study of chemorah basin, Algeria. *Malays. J. Geosci.* 4, 70–78. doi:10.26480/mjg.02.2020.70.78

Lee, S., and Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4, 33–41. doi:10.1007/s10346-006-0047-y

Li, B., Wang, N., and Chen, J. (2021). GIS-based landslide susceptibility mapping using information, frequency ratio, and artificial neural network methods in qinghai province, northwestern China. *Adv. Civ. Eng.* 2021, 1–14. doi:10.1155/2021/4758062

Li, Q., Song, D., Yuan, C., and Nie, W. (2022a). An image recognition method for the deformation area of open-pit rock slopes under variable rainfall. *Meas. (Mahwah. N. J).* 188, 110544. doi:10.1016/j.measurement.2021.110544

Li, Y., Du, L., and Wei, D. (2022b). Multiscale CNN based on component analysis for SAR ATR. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12. doi:10.1109/TGRS.2021.3100137

Majeed, M., Lu, L., Haq, S. M., Waheed, M., Sahito, H. A., Fatima, S., et al. (2022). Spatiotemporal distribution patterns of climbers along an abiotic gradient in jhelum district, Punjab, Pakistan. *Forests* 13, 1244. doi:10.3390/f13081244

Marchesini, I., Mergili, M., Rossi, M., Santangelo, M., Cardinali, M., Ardizzone, F., et al. (2014). "A GIS approach to analysis of deep-seated slope stability in complex geology," in *Landslide science for a safer geoenvironment* (Springer), 483–489.

Milevski, I., Dragićević, S., and Zorn, M. (2019). Statistical and expert-based landslide susceptibility modeling on a national scale applied to North Macedonia. *Open Geosci.* 11, 750–764. doi:10.1515/geo-2019-0059

Miller, S., Brewer, T., and Harris, N. (2009). Rainfall thresholding and susceptibility assessment of rainfall-induced landslides: Application to landslide management in St Thomas, Jamaica. *Bull. Eng. Geol. Environ.* 68, 539–550. doi:10.1007/s10064-009-0232-z

Narimah Samat, M. S. S., and Ismail, M. H. (2020). The integration of gis, ahp, and remote sensing methods for potential areas groundwater: Case study for pontian district, johor, Malaysia. *Malays. J. Geosci.* 5, 06–11. doi:10.26480/mjg.01.2021.06.11

Nolasco-Javier, D., Kumar, L., and Tengonciang, A. M. P. (2015). Rapid appraisal of rainfall threshold and selected landslides in Baguio, Philippines. *Nat. Hazards* 78, 1587–1607. doi:10.1007/s11069-015-1790-y

Oh, H. J., Lee, S., and Hong, S. M. (2017). Landslide susceptibility assessment using frequency ratio technique with iterative random sampling. *J. Sens.* 2017, 1–21. doi:10.1155/2017/3730913

Park, S., Choi, C., Kim, B., and Kim, J. (2013). Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* 68, 1443–1464. doi:10.1007/s12665-012-1842-5

Pham, B. T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P. T., Ngo, T. Q., et al. (2019). A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* 34, 1385–1407. doi:10.1080/10106049.2018.1489422

Pourghasemi, H. R., Moradi, H. R., and Fatemi Aghda, S. M. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* 69, 749–779. doi:10.1007/s11069-013-0728-5

Pourghasemi, H. R., Yansari, T., Panagos, P., and Pradhan, B. (2005). Analysis and evaluation of landslide susceptibility: A review on articlespublished during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arab. J. Geosci.* 11, 193. doi:10.1007/s12517-018-3531-5

Pradhan, B., Chaudhari, A., Adinarayana, J., and Buchroithner, M. F. (2012). Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: A case study at penang Island, Malaysia. *Environ. Monit. Assess.* 184, 715–727. doi:10.1007/s10661-011-1996-8

Pradhan, B. (2010). Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. *Adv. Space Res.* 45, 1244–1256. doi:10.1016/j.asr.2010.01.006

Qasim, M., Hubacek, K., Termansen, M., and Fleskens, L. (2013). Modelling land use change across elevation gradients in district Swat , Pakistan. *Reg. Environ. Change* 13, 567. doi:10.1007/s10113-012-0395-1

Quan, Q., Gao, S., Shang, Y., and Wang, B. (2021). Assessment of the sustainability of Gymnocypris eckloni habitat under river damming in the source region of the Yellow River. *Sci. Total Environ.* 778, 146312. doi:10.1016/j.scitotenv.2021.146312

Raghuvanshi, T. K. (2019). Governing factors influence on rock slope stability – statistical analysis for plane mode of failure. *J. King Saud Univ. - Sci.* 31, 1254–1263. doi:10.1016/j.jksus.2019.01.002

Rahman, G., Rahman, A. U., Bacha, A. S., Mahmood, S., Moazzam, M. F. U., and Lee, B. G. (2020). Assessment of landslide susceptibility using weight of evidence and frequency ratio model in shahpur valley, eastern Hindu Kush. *Nat. Hazards Earth Syst. Sci.*, 1–19. doi:10.21203/rs.3.rs-288102/v1

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth. Sci. Rev.* 180, 60–91. doi:10.1016/j.earscirev.2018.03.001

Ren, Y., Jiang, H., Ji, N., and Yu, H. (2022). Tbsm: A traffic burst-sensitive model for short-term prediction under special events. *Knowl. Based. Syst.* 240, 108120. doi:10.1016/j.knosys.2022.108120

Sadiq Fareed, M. M., Raza, A., Zhao, N., Tariq, A., Younas, F., Ahmed, G., et al. (2022). Predicting divorce prospect using ensemble learning: Support vector machine, linear model, and neural network. *Comput. Intell. Neurosci.* 2022, 1–15. doi:10.1155/2022/3687598

Searle, M. P., Khan, M. A., Fraser, J. E., Gough, S. J., and Jan, M. Q. (1999). The tectonic evolution of the Kohistan-Karakoram collision belt along the Karakoram Highway transect, north Pakistan. *Tectonics* 18, 929–949. doi:10.1029/1999tc900042

Segoni, S., Pappafico, G., Luti, T., and Catani, F. (2020). Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. *Landslides* 17, 2443–2453. doi:10.1007/s10346-019-01340-2

Shah, S. H. I. A., Jianguo, Y., Jahangir, Z., Tariq, A., and Aslam, B. (2022). Integrated geophysical technique for groundwater salinity delineation, an approach to agriculture sustainability for Nankana Sahib Area, Pakistan. *Geomat. Nat. Hazards Risk* 13, 1043–1064. doi:10.1080/19475705.2022.2063077

Shah, S. H. I. A., Yan, J., Ullah, I., Aslam, B., Tariq, A., Zhang, L., et al. (2021). Classification of aquifer vulnerability by using the drastic index and geo-electrical techniques. *WaterSwitzerl.* 13, 2144. doi:10.3390/w13162144

Sharifi, A., Mahdipour, H., Moradi, E., and Tariq, A. (2022). Agricultural field extraction with deep learning algorithm and satellite imagery. *J. Indian Soc. Remote Sens.* 50, 417–423. doi:10.1007/s12524-021-01475-7

Silalahi, F. E. S., PamelaArifianti, Y., and Hidayat, F. (2019). Landslide susceptibility assessment using frequency ratio model in Bogor, West Java, Indonesia. *Geosci. Lett.* 6, 10. doi:10.1186/s40562-019-0140-4

Tahirkheli, R. A. K. (1979). Geology of kohistan and adjoining eurasian and indo-Pakistan continents, Pakistan. *Geol. Bull. Univ. Peshawar* 11, 1–30.

Tariq, A., Mumtaz, F., Zeng, X., Baloch, M. Y. J., and Moazzam, M. F. U. (2022a). Spatio-temporal variation of seasonal heat islands mapping of Pakistan during 2000–2019, using day-time and night-time land surface temperatures MODIS and meteorological stations data. *Remote Sens. Appl. Soc. Environ.* 27, 100779. doi:10.1016/j.rsase.2022.100779

Tariq, A., Riaz, I., Ahmad, Z., Yang, B., Amin, M., Kausar, R., et al. (2020). Land surface temperature relation with normalized satellite indices for the estimation of spatio-temporal trends in temperature among various land use land cover classes of an arid Potohar region using Landsat data. *Environ. Earth Sci.* 79, 40–15. doi:10.1007/s12665-019-8766-2

Tariq, A., and Shu, H. (2020). CA-Markov chain analysis of seasonal land surface temperature and land use landcover change using optical multi-temporal satellite data of Faisalabad, Pakistan. *Remote Sens. (Basel).* 12, 3402–3423. doi:10.3390/rs12203402

Tariq, A., Shu, H., Gagnon, A. S., Li, Q., Mumtaz, F., Hysa, A., et al. (2021a). Assessing burned areas in wildfires and prescribed fires with spectral indices and SAR images in the margalla hills of Pakistan. *Forests* 12, 1371. doi:10.3390/f12101371

Tariq, A., Shu, H., Siddiqui, S., Imran, M., and Farhan, M. (2021b). Monitoring land use and land cover changes using geospatial techniques, a case study of Fateh Jang, Attock, Pakistan. *Geogr. Environ. Sustain.* 14, 41–52. doi:10.24057/2071-9388-2020-117

Tariq, A., Siddiqui, S., Sharifi, A., Hassan, S., and Ahmad, I. (2022b). Impact of spatio - temporal land surface temperature on cropping pattern and land use and land cover changes using satellite imagery , Hafizabad District , Punjab , Province of Pakistan. *Arab. J. Geosci.* 15, 1045–1116. doi:10.1007/s12517-022-10238-8

Tian, H., Huang, N., Niu, Z., Qin, Y., Pei, J., and Wang, J. (2019). Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm. *Remote Sens. (Basel).* 11, 820–823. doi:10.3390/rs11070820

Tian, H., Pei, J., Huang, J., Li, X., Wang, J., Zhou, B., et al. (2020). Garlic and winter wheat identification based on active and passive satellite imagery and the Google Earth engine in northern China. *Remote Sens. (Basel).* 12, 3539–3617. doi:10.3390/rs12213539

Tian, H., Qin, Y., Niu, Z., Wang, L., and Ge, S. (2021a). Summer maize mapping by compositing time series sentinel-1A imagery based on crop growth cycles. *J. Indian Soc. Remote Sens.* 49, 2863–2874. doi:10.1007/s12524-021-01428-0

Tian, H., Wang, Y., Chen, T., Zhang, L., and Qin, Y. (2021b). Early-season mapping of winter crops using sentinel-2 optical imagery. *Remote Sens. (Basel).* 13, 3822–3911. doi:10.3390/rs13193822

Tosic, R., Dragicevic, S., Zorn, M., and Lovric, N. (2014). Landslide susceptibility zonation: A case study of the municipality of banja luka (Bosnia and Herzegovina). *ACTA Geogr. Slov. Zb.* 54, 190–202. doi:10.3986/ags54307

Ullah, I., Aslam, B., Shah, S. H. I. A., Tariq, A., Qin, S., Majeed, M., et al. (2022). An integrated approach of machine learning, remote sensing, and GIS data for the landslide susceptibility mapping. *Land* 11, 1265. doi:10.3390/land11081265

Vakhshoori, V., and Zare, M. (2016). Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomat. Nat. Hazards Risk* 7, 1731–1752. doi:10.1080/19475705.2016.1144655

Wahla, S. S., Kazmi, J. H., Sharifi, A., Shirazi, S. A., Tariq, A., and Joyell Smith, H. (2022). Assessing spatio-temporal mapping and monitoring of climatic variability using SPEI and RF machine learning models. *Geocarto Int.* 0, 1–20. doi:10.1080/10106049.2022.2093411

Wang, P., Wang, L., Leung, H., and Zhang, G. (2021a). Super-resolution mapping based on spatial–spectral correlation for spectral imagery. *IEEE Trans. Geosci. Remote Sens.* 59, 2256–2268. doi:10.1109/tgrs.2020.3004353

Wang, Q., Zhou, G., Song, R., Xie, Y., Luo, M., and Yue, T. (2022). Continuous space ant colony algorithm for automatic selection of orthophoto mosaic seamline network. *ISPRS J. Photogramm. Remote Sens.* 186, 201–217. doi:10.1016/j.isprsjprs.2022.02.011

Wang, S., Zhang, K., Chao, L., Li, D., Tian, X., Bao, H., et al. (2021b). Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards. *J. Hydrol. X.* 603, 126964. doi:10.1016/j.jhydrol.2021.126964

Waqas, H., Lu, L., Tariq, A., Li, Q., Baqa, M. F., Xing, J., et al. (2021). *Flash flood susceptibility assessment and zonation using an integrating analytic hierarchy process and frequency ratio model for the chitral district, khyber pakhtunkhwa*, 13. pakistan. doi:10.3390/w13121650*WaterSwitzerl.*

Wu, C. H., and Chen, S. C. (2009). Determining landslide susceptibility in Central Taiwan from rainfall and six site factors using the analytical hierarchy process method. *Geomorphology* 112, 190–204. doi:10.1016/j.geomorph.2009.06.002

Xie, W., Li, X., Jian, W., Yang, Y., Liu, H., Robledo, L. F., et al. (2021a). A novel hybrid method for landslide susceptibility mapping-based geodetector and machine learning cluster: A case of xiaojin county, China. *ISPRS Int. J. Geoinf.* 10, 93. doi:10.3390/ijgi10020093

Xie, W., Nie, W., Saffari, P., Robledo, L. F., Descote, P. Y., and Jian, W. (2021b). Landslide hazard assessment based on Bayesian optimization–support vector machine in Nanping City, China. *Nat. Hazards* 109, 931–948. doi:10.1007/s11069-021-04862-y

Xu, C., Shyu, J. B. H., and Xu, X. (2014). Landslides triggered by the 12 january 2010 port-au-prince, Haiti, <i>M</i>_w = 7.0 earthquake: Visual interpretation, inventory compiling, and spatial distribution statistical analysis. *Nat. Hazards Earth Syst. Sci.* 14, 1789–1818. doi:10.5194/nhess-14-1789-2014

Yin, L., Wang, L., Keim, B. D., Konsoer, K., and Zheng, W. (2022a). Wavelet analysis of dam injection and discharge in three gorges dam and reservoir with precipitation and river discharge. *WaterSwitzerl.* 14, 567. doi:10.3390/w14040567

Yin, L., Wang, L., Zheng, W., Ge, L., Tian, J., Liu, Y., et al. (2022b). Evaluation of empirical atmospheric models using swarm-C satellite data. *Atmos. (Basel)* 13, 294–315. doi:10.3390/atmos13020294

Yue, Z., Zhou, W., and Li, T. (2021). Impact of the Indian ocean dipole on evolution of the subsequent ENSO: Relative roles of dynamic and thermodynamic processes. *J. Clim.* 34, 3591–3607. doi:10.1175/JCLI-D-20-0487.1

Zamani, A., Sharifi, A., Felegari, S., Tariq, A., and Zhao, N. (2022). Agro climatic zoning of saffron culture in miyaneh city by using WLC method and remote sensing data. *Agriculture* 12, 118–215. doi:10.3390/agriculture12010118

Zhan, C., Dai, Z., Samper, J., Yin, S., Ershadnia, R., Zhang, X., et al. (2022). An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network. *J. Hydrol. X.* 610, 127844. doi:10.1016/j.jhydrol.2022.127844

Zhang, K., Ali, A., Antonarakis, A., Moghaddam, M., Saatchi, S., Tabatabaeenejad, A., et al. (2019a). The sensitivity of north American terrestrial carbon fluxes to spatial and temporal variation in soil moisture: An analysis using radar-derived estimates of root-zone soil moisture. *J. Geophys. Res. Biogeosci.* 124, 3208–3231. doi:10.1029/2018JG004589

Zhang, K., Wang, S., Bao, H., and Zhao, X. (2019b). Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China. *Nat. Hazards Earth Syst. Sci.* 19, 93–105. doi:10.5194/nhess-19-93-2019

Zhang, X., Ma, F., Yin, S., Wallace, C. D., Soltanian, M. R., Dai, Z., et al. (2021). Application of upscaling methods for fluid flow and mass transport in multi-scale heterogeneous media: A critical review. *Appl. Energy* 303, 117603. doi:10.1016/j.apenergy.2021.117603

Zhang, Z., Luo, C., and Zhao, Z. (2020). Application of probabilistic method in maximum tsunami height prediction considering stochastic seabed topography. *Nat. Hazards* 104, 2511–2530. doi:10.1007/s11069-020-04283-3

Zhao, F., Zhang, S., Du, Q., Ding, J., Luan, G., and Xie, Z. (2021). Assessment of the sustainable development of rural minority settlements based on multidimensional data and geographical detector method: A case study in dehong, China. *Socioecon. Plann. Sci.* 78, 101066. doi:10.1016/j.seps.2021.101066

Zhou, G., Long, S., Xu, J., Zhou, X., Song, B., Deng, R., et al. (2021a). Comparison analysis of five waveform decomposition algorithms for the airborne LiDAR echo signal. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 7869–7880. doi:10.1109/JSTARS.2021.3096197

Zhou, G., Zhang, R., and Huang, S. (2021b). Generalized buffering algorithm. *IEEE Access* 9, 27140–27157. doi:10.1109/ACCESS.2021.3057719

Zhou, W., Lv, Y., Lei, J., and Yu, L. (2021c). Global and local-contrast guides content-aware fusion for RGB-D saliency prediction. *IEEE Trans. Syst. Man. Cybern. Syst.* 51, 3641–3649. doi:10.1109/tsmc.2019.2957386

Zhu, B., Zhong, Q., Chen, Y., Liao, S., Li, Z., Shi, K., et al. (2022a). A novel reconstruction method for temperature distribution measurement based on ultrasonic tomography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 69, 2352–2370. doi:10.1109/TUFFC.2022.3177469

Zhu, Z., Wu, Y., and Liang, Z. (2022b). Mining-induced stress and ground pressure behavior characteristics in mining a thick coal seam with hard roofs. *Front. Earth Sci.* 10, 1–12. doi:10.3389/feart.2022.843191

Zuhairi, A., Nur Syahira Azlyn, A., Nur Suhaila, M. R., and . Mohd Zaini, M. (2020). Land use classification and mapping using Landsat imagery for gis database in langkawi Island. *Sci. Herit. J.* 4, 59–63. doi:10.26480/gws.02.2020.59.63

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/feart.2022.1057796

OPEN ACCESS

EDITED BY
Wei Zhao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Veronica Pazzi,
University of Trieste, Italy
Ionut Cristi Nicu,
Norwegian Institute for Cultural Heritage Research, Norway

*CORRESPONDENCE
Nishan Kumar Biswas,
n.biswas@nasa.gov

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 30 September 2022
ACCEPTED 21 October 2022
PUBLISHED 09 November 2022

CITATION
Biswas NK, Stanley TA, Kirschbaum DB, Amatya PM, Meechaiya C, Poortinga A and Towashiraporn P (2022), A dynamic landslide hazard monitoring framework for the Lower Mekong Region. *Front. Earth Sci.* 10:1057796. doi: 10.3389/feart.2022.1057796

A dynamic landslide hazard monitoring framework for the Lower Mekong Region

Nishan Kumar Biswas[1,2,3]*, Thomas A. Stanley[1,2,3], Dalia B. Kirschbaum[4], Pukar M. Amatya[1,2,3], Chinaporn Meechaiya[5], Ate Poortinga[6] and Peeranan Towashiraporn[5]

[1]University of Maryland, Baltimore County, Baltimore, MD, United States, [2]Goddard Earth Sciences Technology and Research II, Baltimore, MD, United States, [3]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, United States, [4]Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, United States, [5]Asian Disaster Preparedness Center, Bangkok, Thailand, [6]Spatial Informatics Group, LLC, Pleasanton, CA, United States

The Lower Mekong region is one of the most landslide-prone areas of the world. Despite the need for dynamic characterization of landslide hazard zones within the region, it is largely understudied for several reasons. Dynamic and integrated understanding of landslide processes requires landslide inventories across the region, which have not been available previously. Computational limitations also hamper regional landslide hazard assessment, including accessing and processing remotely sensed information. Finally, open-source software and modelling packages are required to address regional landslide hazard analysis. Leveraging an open-source data-driven global Landslide Hazard Assessment for Situational Awareness model framework, this study develops a region-specific dynamic landslide hazard system leveraging satellite-based Earth observation data to assess landslide hazards across the lower Mekong region. A set of landslide inventories were prepared from high-resolution optical imagery using advanced image-processing techniques. Several static and dynamic explanatory variables (i.e., rainfall, soil moisture, slope, relief, distance to roads, distance to faults, distance to rivers) were considered during the model development phase. An extreme gradient boosting decision tree model was trained for the monsoon period of 2015–2019 and the model was evaluated with independent inventory information for the 2020 monsoon period. The model performance demonstrated considerable skill using receiver operating characteristic curve statistics, with Area Under the Curve values exceeding 0.95. The model architecture was designed to use near-real-time data, and it can be implemented in a cloud computing environment (i.e., Google Cloud Platform) for the routine assessment of landslide hazards in the Lower Mekong region. This work was developed in collaboration with scientists at the Asian Disaster Preparedness Center as part of the NASA SERVIR Program's Mekong hub. The goal of this work is to develop a suite of tools and services on accessible open-source platforms that support and enable stakeholder communities to better assess landslide hazard and exposure at local to regional scales for decision making and planning.

KEYWORDS

Landslides, lower mekong countries, google earth engine (GEE), XGBoost (extreme gradient boosting), machine learning, satellite earth observation, LEWS

1 Introduction

Landslides are one of the most dangerous natural hazards that can cause a sudden loss of human life and substantial property damage. The Lower Mekong Region (LMR) is highly susceptible to landslides (Figure 1) due to steep terrain combined with intense rainfall (Stanley and Kirschbaum, 2017). From the inset plot of the Figure 1, four out of five countries (Cambodia, Myanmar, Vietnam, and Thailand) have well covered medium to high susceptible topography with more than 50 fatalities within 2015–2020.

For the past few decades, several studies have been performed with various approaches to quantify landslide susceptibility in the LMR region at a mostly regional scale. Those approaches can be classified into different types: 1) Machine learning, deep learning and Artificial Intelligence based methods; 2) Simple GIS analysis-based methods; and 3) Physical model-based methods. In Table 1, a summary of the relevant studies in the different categories is shown. Among the latest methods, Tran et al. (2020) used a machine-learning algorithm to characterize the landslide hazards in the Nam Dam Commune, Vietnam region. Van Dao et al. (2020) found that spatially explicit deep learning (DL) neural network models performed better than other models for the prediction of landslide susceptibility in the Muong Lay district (Vietnam). Bui et al. (2019) compared Deep Learning Neural Network (DLNN) with four different types of machine learning models to produce landslide susceptibility. While all these studies were performed to characterize the landslide susceptibility in the LMR using different approaches, to our best knowledge, no study focused on dynamic landslide hazard assessment in an integrated way for the whole region. A few possible reasons for not having a near-real-time integrated

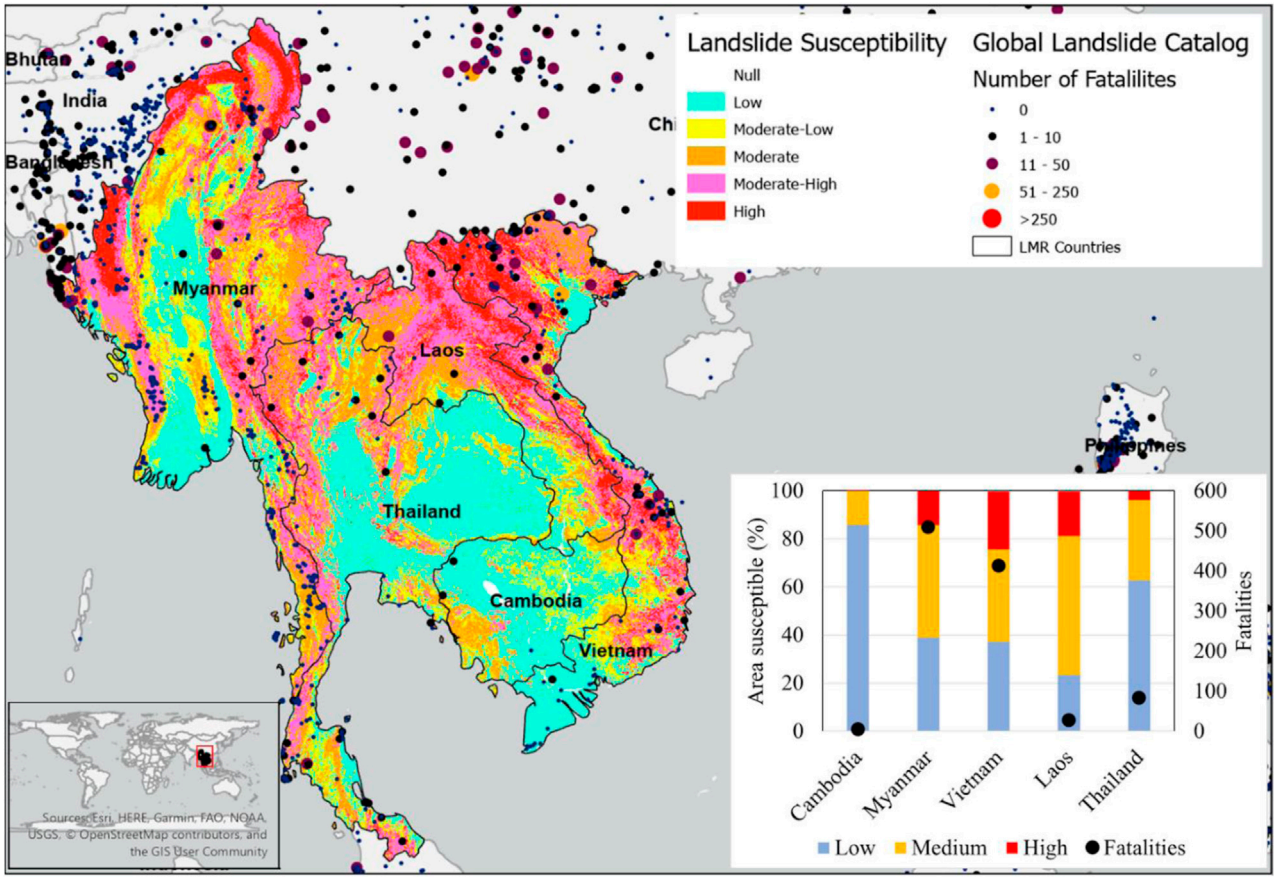


FIGURE 1
Distribution of reported landslide fatalities across the Lower Mekong Region based on NASA's Global Landslide Catalog (Kirschbaum et al., 2015), with NASA's global landslide susceptibility map (Stanley and Kirschbaum, 2017).

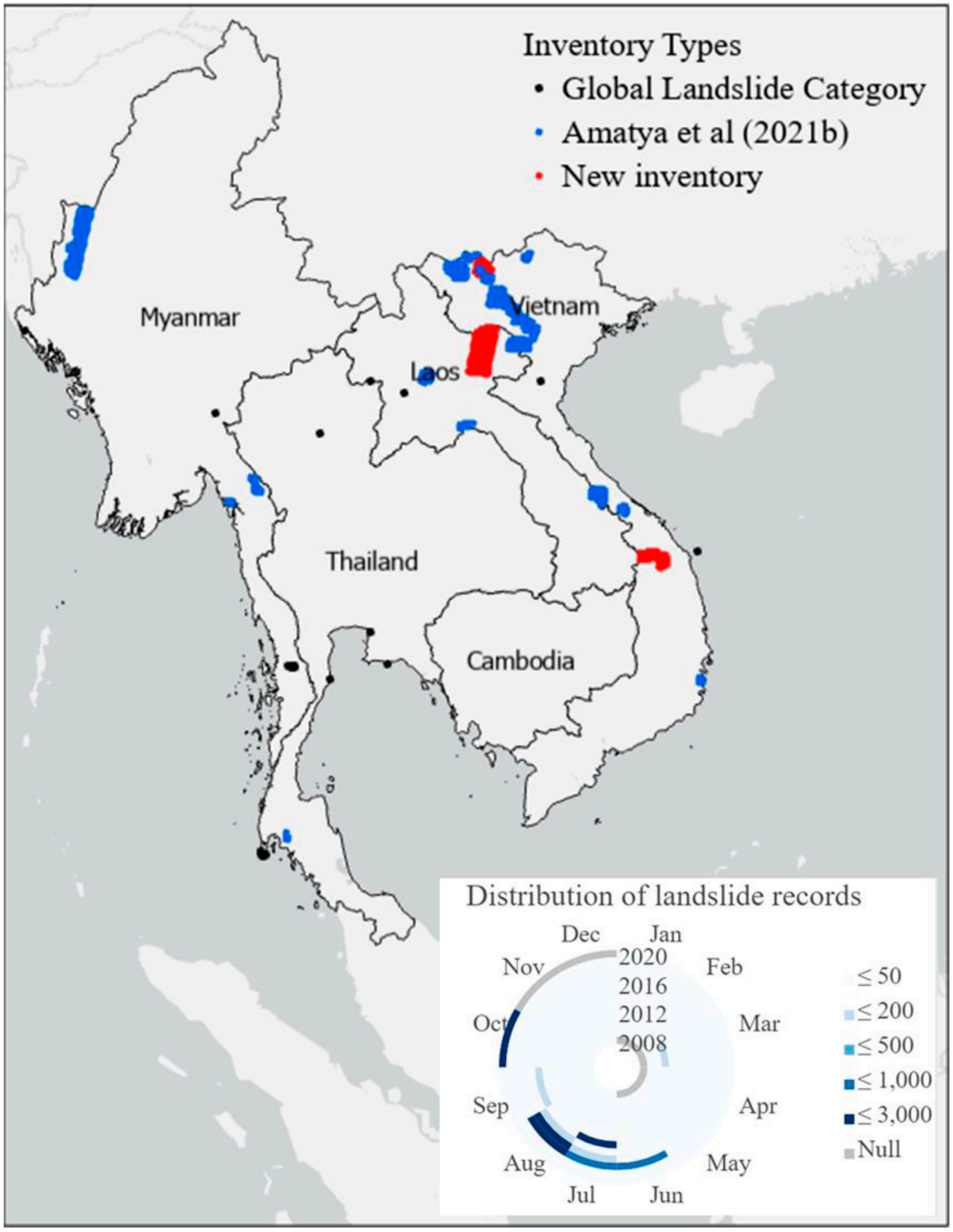


FIGURE 2
Landslide inventory records generated by SaLaD-CD Algorithm over the LMR.

TABLE 1 Geographic region, landslide relevant studies, types, and major findings.

Country	Study	Model	Comments
Cambodia	Lee at al. (2006)	Frequency ratio and logistic regression	Similar prediction accuracy
Laos	Hearn and Pongpanya (2021)	A simple index of vulnerability assessment	Only for the road network
Myanmar	Dang et al. (2020)	Deterministic models	Velocity, volume and locations
	Kolomazník et al. (2021)	Provided evidence-based support to the design of the road corridor	landslide inventory susceptibility map prepared
	Naing et al. (2022)	GIS analysis	Causes of landslides studied for topographic variables
	Panday and Dong (2021)	Investigative study	Relationship between landslides and topographic features was studied
	Titti et al. (2021)	Landslide susceptibility zoning using weight of evidence method	Mapped the landslide susceptibility
Thailand	Asian Disaster Preparedness Center et al. (2008)	Wide variety of assessments at the city and slope scales	Combines critical API map with automated rain gauges for real-time warnings
	Intarawichian & Dasananda (2011)	Frequency Ratio model	Susceptibility maps were prepared
	Jotisankasa and Vathananukij (2008)	1D infiltration modeling and slope stability analysis	300 mm rainfall over 1–4 days may trigger debris flows
	Kanjanakul et al. (2016)	Combines SEEP/W and SLOPE/W	Calculated a critical API of 380 mm
	Komori et al. (2018)	Logistic regression and Richard's equation	Landslide hazard will increase due to climate change
	Nawagamuwa et al. (2013)	Analysis of historical landslides in several countries	Thai landslides caused by rainfall over 3 days
	Oh et al. (2009)	Frequency ratio and logistic regression	Frequency ratio performed better
	Ohtsu et al. (2012)	Modified API, 10-min rainfall data	Antecedent rainfall is significant for 2 days prior to event
	Ono et al. (2014)	Infinite slope stability model	Model was not sensitive to soil parameters
	Phien-Wej et al. (1993)	Retrospective analysis of the 1988 landslide disaster	Most failures in weathered granite; 475 mm rainfall/day
	Salee et al. (2022)	Event rainfall depth-duration threshold	Best results when combined with 20-day cumulative rainfall
	Schmidt-Thomé et al. (2018)	Slope threshold of 30°, combined with "loose soil"	Rainfall thresholds by susceptibility zone imply use in a real-time system
	Sidle & Ziegler, (2017)	Intensity-duration (ID) threshold, combined with 2-day API	Forest canopies have little effect on landslides in this area
	Soralump, (2010)	Critical API based on rock groups	Applied both locally and nationally
Vietnam	Bui et al. (2017)	Compared least squares support vector machines (LSSVM) to other methods	LSSVM Performed best
	Bui et al. (2012)	Support vector machines (SVM), decision tree (DT), and Naïve Bayes (NB)	SVM performed best among all
	Bui et al. (2011)	statistical index and the logistic regression	Performed same
	Tran et al. (2020)	Hyperpipes (HP) algorithm AdaBoost (AB), Bagging (B), Dagging, Decorate, and Real AdaBoost (RAB)	AdaBoost performed best
	Truong et al. (2018)	BE-LMtree and Support Vector Machine	BE-LMTree model performs better
	Van Dao et al. (2020)	Deep learning compared with other models: quadratic discriminant analysis, Fisher's linear discriminant analysis, and multi-layer perceptron neural network	Deep learning performed best

hazard system include: 1) dependability of a specific model/algorithm to produce reliable landslide hazard with spatio-temporal consistency; 2) absence of sufficient landslide inventories; 3) lack of access to or availability of relevant hydrometeorological parameters that are either satellite-based or *in situ*; 4) limitations associated with the computational resources.

In this study, we focused on minimizing the above-mentioned key issues that hindered near-real-time landslide hazard assessment. To address the first issue, we used a regionalized version of the global Landslide Hazard Assessment for Situational Awareness model (LHASA) (Stanley et al., 2021). LHASA was developed to provide an indication of where and when landslides may be likely around the world. This model uses surface susceptibility (including slope, vegetation, road networks, geology, and forest cover loss) and satellite rainfall data from the Global Precipitation Measurement (GPM) mission to provide near-real-time hazards. The use of

TABLE 2 Static variables and the data sources used in LHASA-Mekong.

Variable	Source
Distance to faults	GEM Global Active Faults (GEM Hazard Team, 2019)
Lithology Strength	Global lithological map (Hartmann and Moosdorf, 2012)
Slope	SRTM DEM
Relief	
Topographic Wetness Index (TWI)	
Distance to Rivers	HydroRIVERS (USGS HydroSHEDS)
Distance to Roads	Global Roads Inventory Project (Meijer et al., 2018)

TABLE 3 LHASA-Mekong model description with the interaction and monotone constraints.

Model parameters	Tuned value
Maximum depth	2
Eta	0.05
Objective	binary: logistic
Tree method	Exact
Maximum Delta Step	1
Subsample	0.5
Monotone Constraints	Distance to Roads, Distance to Faults, Distance to rivers is an inverse relationship, all others are having linear relationship
Interaction Constraints	Precipitation can interact with others

machine learning, when combined with other improvements to LHASA, doubled the system's ability, relative to the prior version (Stanley and Kirschbaum, 2017). Based on this experience, a regional version of the LHASA model was proposed in this study. To address the dearth of published landslide inventories, Amatya et al. (2021b) developed a change detection-based Semi-Automatic Landslide Detection (SALaD-CD) system to detect landslides in the LMR. To address the regional landslide triggering topographic and meteorological data scarcity issue, we used NASA Earth observation datasets, which over the last few decades have demonstrated potential to be used in conjunction with ground-based observations for monitoring water resources and hydroclimatic disasters. Finally, the emergence of cloud computing and advanced IT infrastructure reduces the need for local computational resources, which is helpful for many stakeholder organizations (Biswas and Hossain, 2018; Biswas et al., 2021; Biswas et al., 2020). The freely available framework, Google Earth Engine-GEE (Gorelick et al., 2017), has been widely used for cloud-based scientific analysis and visualization of geospatial datasets by academic, non-profit, business, and government users. It hosts a public data archive that includes a record of more than 4 decades and is increasing every day. This framework has been helping stakeholders and decision makers use a plethora of satellite data without incurring any additional charges and widens the possibilities for developing nations. By integrating the Google Cloud Platform (GCP) with GEE, a regional version of LHASA named LHASA-Mekong was trained, validated, and implemented in this study.

Using LHASA-Mekong, the key research questions addressed in this study are: 1) What level of accuracy can be attained in assessing landslide hazard activity at a regional scale using globally available hydrometeorological and topographical data? and 2) How can a disaster modeling framework be implemented in a cloud environment for near-real-time prediction? In the paper, relevant datasets and methodology of the study are discussed in Section 2, Section 3 emphasizes the results, major findings, and discussion followed by the conclusions and future scope of the study.

2 Data and methodology

2.1 Data

2.1.1 Landslide inventory preparation

The landslide inventories used in this research were mapped utilizing high-resolution satellite imagery from Planet (Planet Team, 2017) using the modified framework of Semi-Automatic Landslide Detection (SALaD) system (Amatya et al., 2021a). The SALaD uses object-based image analysis and machine learning to map landslides. A change detection-based approach was introduced to the SALaD framework (SALaD-CD) utilizing pre- and post-event imagery (Amatya et al., 2021b). The new framework incorporates image normalization, image co-registration, and change detection. In addition to the 22 inventories (2 in Laos, 4 in Myanmar, 1 in Thailand, and 15 in Vietnam) released by Amatya et al., 2021 using SALaD-CD, three new inventories were added in this study (Xam Neua, Laos; Quang Nam, Vietnam; and Sapa, Vietnam). 112 landslides recorded in the Global Landslide Catalog (Kirschbaum et al., 2015) were also included. The map of available landslide inventories and the monthly distribution of these data are shown in Figure 2A.

2.1.2 Satellite precipitation estimates

To fill the ground-based observation gap, we depended on satellite-based observations that show reasonable accuracy when compared to the ground-based data. For the precipitation product, we used the Global Precipitation Measurement (GPM) satellite provided Integrated Multi-satellitE Retrievals for GPM (IMERG) version 6 product (Huffman et al., 2015). The IMERG rainfall product was developed using a unified

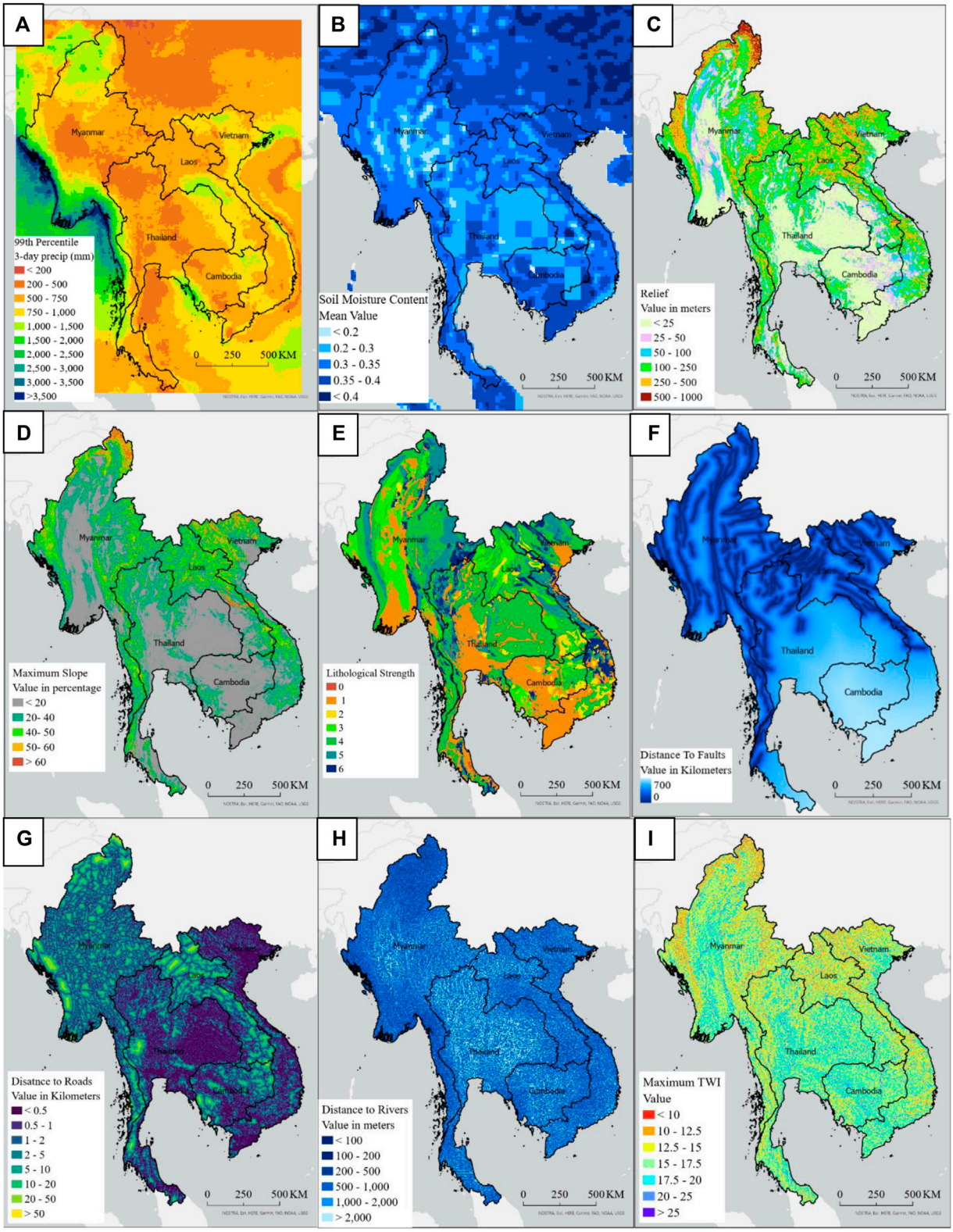


FIGURE 3
Explanatory variables used during the development of LHASA-Mekong Model, **(A)** 99th percentile of 3-day accumulated precipitation, **(B)** Mean of CFS soil moisture variable, **(C)** Relief, **(D)** Slope, **(E)** Lithological strength, **(F)** Distance to faults, **(G)** Distance to roads, **(H)** Distance to rivers, and **(I)** Maximum topographic wetness index (TWI).

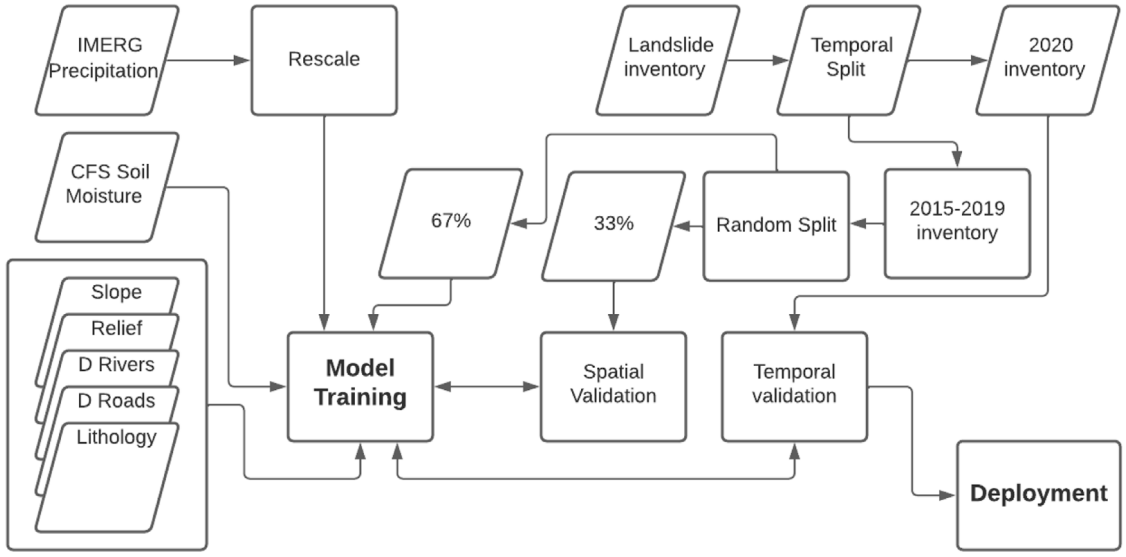


FIGURE 4
Methodology of the development of LHASA-Mekong model.

algorithm that provides rainfall estimates by combining data from microwave and infrared instruments on satellites along with precipitation gauge data for calibration. The native spatial resolution of the data is 0.1° by 0.1° (approximately a 10-km by 10-km grid) with a half-hourly temporal resolution. The half-hourly data was converted into 3-day accumulations at the per-pixel basis to feed the model. The mean of 3-day accumulated precipitation is shown in Figure 3B and the 99th percentile of 3-day accumulated precipitation is shown in Figure 3C.

2.1.3 Climate forecast system soil moisture

The Climate Forecast System (CFS) version 2 was developed at the Environmental Modeling Center at the National Centers for Environmental Prediction (NCEP) (Saha et al., 2011). It is a fully coupled model representing the interaction between the Earth's atmosphere, oceans, land, and sea-ice and became operational at NCEP in March 2011. This dataset was chosen because of its acceptability and availability in GEE with the lowest lag-time (earth engine asset id: NOAA/CFSV2/FOR6H). Minimized lag-time and consistent data availability is important for any kind of operational product. The spatial resolution of the data is 0.2° and the variable used in this study was "Volumetric_Soil_Moisture_Content_depth _below_surface_layer_150_cm." A snapshot of the spatial variability of soil moisture data is shown in Figure 3B.

2.1.4 Static variables

Besides the dynamic variables, several static variables (such as slope, relief, distance to rivers, distance to roads, and lithological strength) were also included in the model. Among the static variables, slope and relief were calculated from the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model (DEM). The native resolution of the SRTM DEM is 30 m. Lithological strength data were leveraged from the global LHASA model as that was the only consistent and reliable data related to lithological strength for this region. Spatial variability of these parameters is shown in Figure 3. Table 2 summarizes the static variables, and the data sources are mentioned.

2.2 Methodology

The overall methodology of the model development is described in Figure 4.

2.2.1 Rescaling of precipitation

To reflect the climate of each location, we normalized the IMERG 3-day accumulated precipitation data using the 99th percentile of the historical 3-day accumulated precipitation. For each of the pixels within the study area, a log-normal distribution was fitted to the 99th percentile. This was calculated by fitting a log-normal distribution for each pixel from daily rainfall records spanning 2000–2020, then calculating the 99th percentile from the fitted distribution. Finally, this precipitation was used to rescale the 3-day accumulated precipitation and the normalized precipitation was fed into the model framework. A map of the 3-day accumulated precipitation is shown in Figure 3A.

2.2.2 Exploratory data analysis

An exploratory data analysis was performed before training the model. Correlation between some of the most important

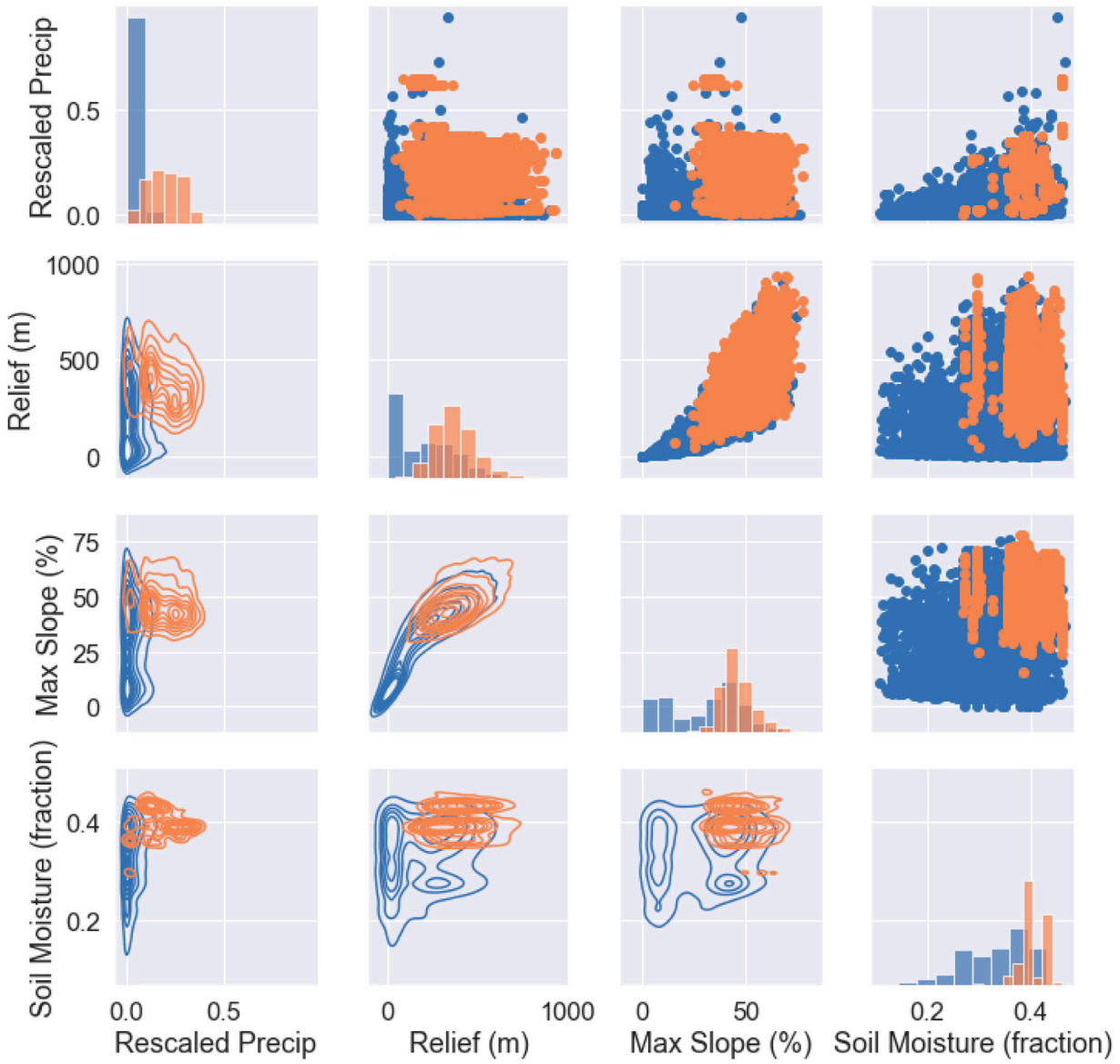


FIGURE 5
Distributions of the most important variables for landslides (orange) and non-landslide grid cells (blue). The highest correlation (0.45) was between maximum slope and relief.

factors that drive landslide initiation such as precipitation, slope, relief, and soil moisture over the landslide and non-landslide grids is shown in Figure 5.

2.2.3 Model training and validation

Two dynamic variables (precipitation and soil moisture) and five static variables were used to train the model. The landslide inventory was divided into two parts, the 2015–2019 inventory was used to train the model and the 2020 inventory was used for the temporal validation of the model. The 2015–2019 inventory was further split randomly into a part for training the model (67%) and for validating the model (33%). The XGBoost python library was used for training of the model. Based on the global landslide modelling experience, we found that XGBoost model has unique features such as interaction and monotone constraints which makes it more appropriate for landslide hazard quantification (Stanley et al., 2021). Interested users are suggested to see Stanley et al. (2021) and https://xgboost.readthedocs.io/en/stable/to know about the XGBoost model. Model configuration parameters are shown in Table 3. The receiver operating characteristics (ROC) curve was used to quantify the accuracy of the model. In order to generate binary statistics, a probability threshold of 0.08 was chosen based on the 99th percentile of the probability distribution of the non-landslide sample.

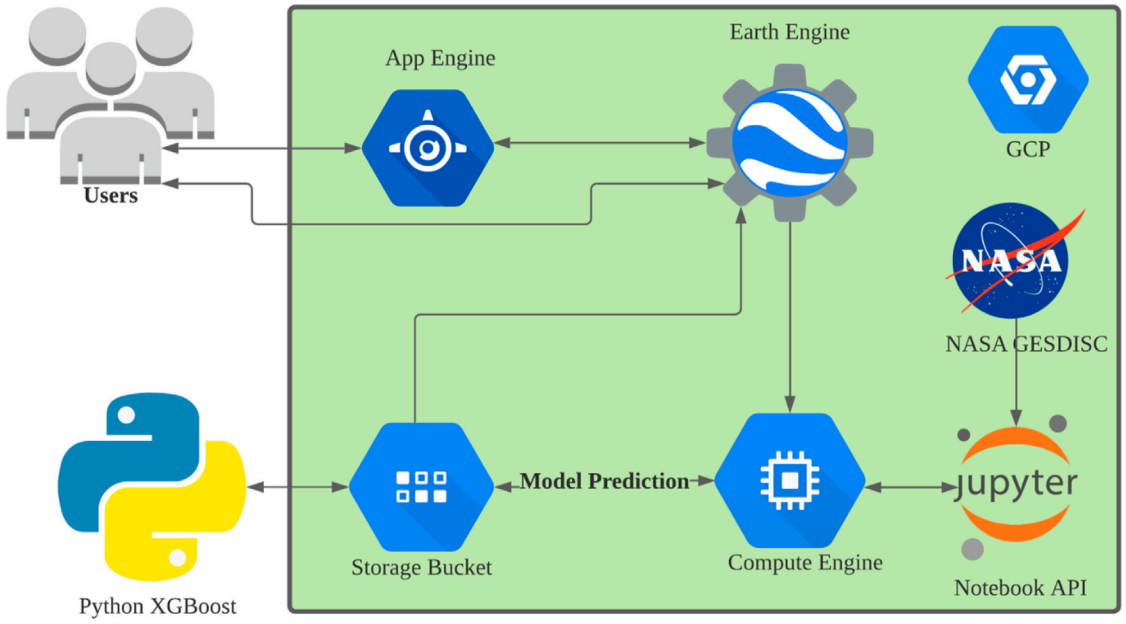


FIGURE 6
LHASA-Mekong Model implementation in Google Cloud Platform (GCP).

2.2.4 Python package with google earth engine for operationalization

Upon finalizing the model, it was implemented in Google Cloud Platform (GCP) for the operational prediction. An illustration on implementation of the LHASA-Mekong model in GCP environment is shown in Figure 6. The finalized version of the model was hosted in GCP bucket storage along with the necessary metadata and configuration information. Compute Engine and Notebook API were used together to deploy the model. It was found that the latest IMERG early run precipitation data was not available in GEE. Thus, the notebook API script was used to pull data from the Goddard Earth Sciences Data and Information Services Center server of NASA Goddard Space Flight Center (GSFC). Other dynamic variables (i.e., IMERG late run precipitation data and CFS soil moisture) were captured from GEE. Then the model was imported to make the prediction and the final hazard probability map was pushed back to GCP bucket storage. Finally, the hazard probability map was pushed to GEE as assets from the GCP bucket storage. The hazard probability data can be accessed through GEE and through API services such as Google App Engine.

3 Results and discussion

3.1 Model training and validation

The Area Under the Curve (AUC) value was 99% in the spatial validation of the model using a random 33% split of the sample data. During the temporal validation of the model, the AUC values were similar to that of the spatial validation. Left panel of Figure 7 shows a feature importance plot combining three different metrics (gain, cover, and weight). Gain is the most relevant attribute to interpret the relative importance of each feature. Cover measures the relative quantity of observations concerned by a feature. Most of the gain was attained by the precipitation and the second-most important variable was maximum slope. Relief also showed similar importance compared to the maximum slope. Soil moisture was incorporated in the model as a dynamic variable, but its importance was quite low compared to the maximum slope and relief. This can be explained by the lower variability and lower accuracy of the variable. The SHapley Additive exPlanations (SHAP) package from python (link: https://shap.readthedocs.io/en/latest/index.html) was also used to explain the feature importance plot of the model, which is shown in right panel of Figure 7. Here it can be seen that precipitation, soil moisture, maximum slope, and relief have a higher positive impact on the predictability of the model. Distance to roads and distance to faults have more negative impacts on the model. Some of the precipitation and soil moisture also have a negative impact on the model due to the frequent occurrence of very moist conditions in flat regions of the LMR.

3.2 Event-specific assessment

In addition to standard validation metrics, we quantified the model accuracy over six large landslide events. Three events were

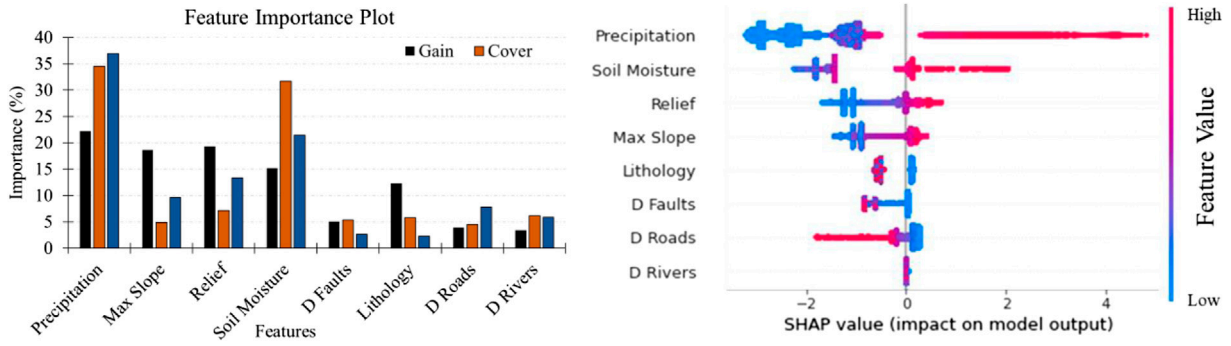


FIGURE 7
Feature importance plot of the LHASA-Mekong model (left) and SHAP value plot (right). Precipitation had the greatest effect on model outcomes, primarily raising hazard values.

selected in the different regions from the training period of 2015–2019, and three of the events were picked from the validation period. The news sources for all these events can be found in Amatya et al. (2021). The areas selected from the model training period are: 1) the Hakha-Falam area of Myanmar where a number of landslide events happened on 30 July 2015; 2) the Muong La, Vietnam area for the date of 3 August 2017; and 3) the Xieng Ngeun, Laos area for the date of 31 August 2018. It was found that for the first two events, the true positive rate of the model was higher than 99%. For the third event, it was found that IMERG estimated rainfall underestimated the ground-measured amount of rainfall provided by Mekong River Commission (MRC) and thus the model-generated hazard missed some of the landslide events, reducing the true positive rate to 97%. The model was further validated for the three individual events in 2020 that happened in the Central Vietnam region. Almost all the landslide locations were identified as potentially hazardous (p>0.08), resulting in a true positive rate (TPR) above 90%. Spatial maps of the predictions for all events show a more extensive zone of a hazard than the area known to be affected by landslides (Figure 8).

3.3 Retrospective simulation and seasonality analysis

To better understand the seasonal and geographic distribution of landslide hazards across the LMR, the model was run retrospectively for the period 2015–2020. The long-term hazard probability map (Figure 9B) was qualitatively compared with the global landslide susceptibility map (Figure 9A) prepared by Stanley and Kirschbaum (2017) and both maps match quite well, indicating that LHASA-Mekong may correctly reflect the geographic distribution of landslide hazard which is coherence to the locations with higher topographic slope, high amount of rainfall and low lithological strength regions.

The seasonal variation of landslide hazard probabilities is shown in Figures 9C–F. From Figure 7C, the whole LMR had a very low chance of having landslide probability, due to extremely low rainfall that occurs in the dry season. The April-May-June map shows higher probabilities in the western zone of LMR due to the comparatively early arrival of the tropical monsoon period. The same pattern can also be seen in Figures 9E,F with 7f showing very high hazard in the Central Vietnam region due to the late arrival of the monsoon. For the same reason, the Myanmar region showed a very low hazard probability during the same period of the year (October-November-December). When the mean hazard probability was converted into a time-series mean, the seasonality pattern was clearer. In Figure 10, a time-series comparison of weekly precipitation and weekly mean hazard probability (along with range) was shown. From Figure 10, the more frequent occurrence of higher landslide probability corresponds to a higher amount of rainfall. There were a few higher probabilities occurring in the late monsoon and dry season and this was due to the higher rainfall reported by the IMERG precipitation.

3.4 Discussion

The LHASA-Mekong model showed an AUC of 99%, which is sometimes treated as overfitting. To reduce overfitting, we used monotone constraints and interaction constraints that are helpful in reducing this nature of machine learning models. In order to re-validate the applicability of the model, we took the following steps: 1) independently validated the model with the 2020 landslide inventory that was not included in the training sample, 2) checked the model prediction accuracy against eight inventories spanning from 2015–2020, 3) evaluated the predicted hazard outputs at annual and seasonal scale to make sure that the model has reliable and consistent accuracy temporally and spatially.

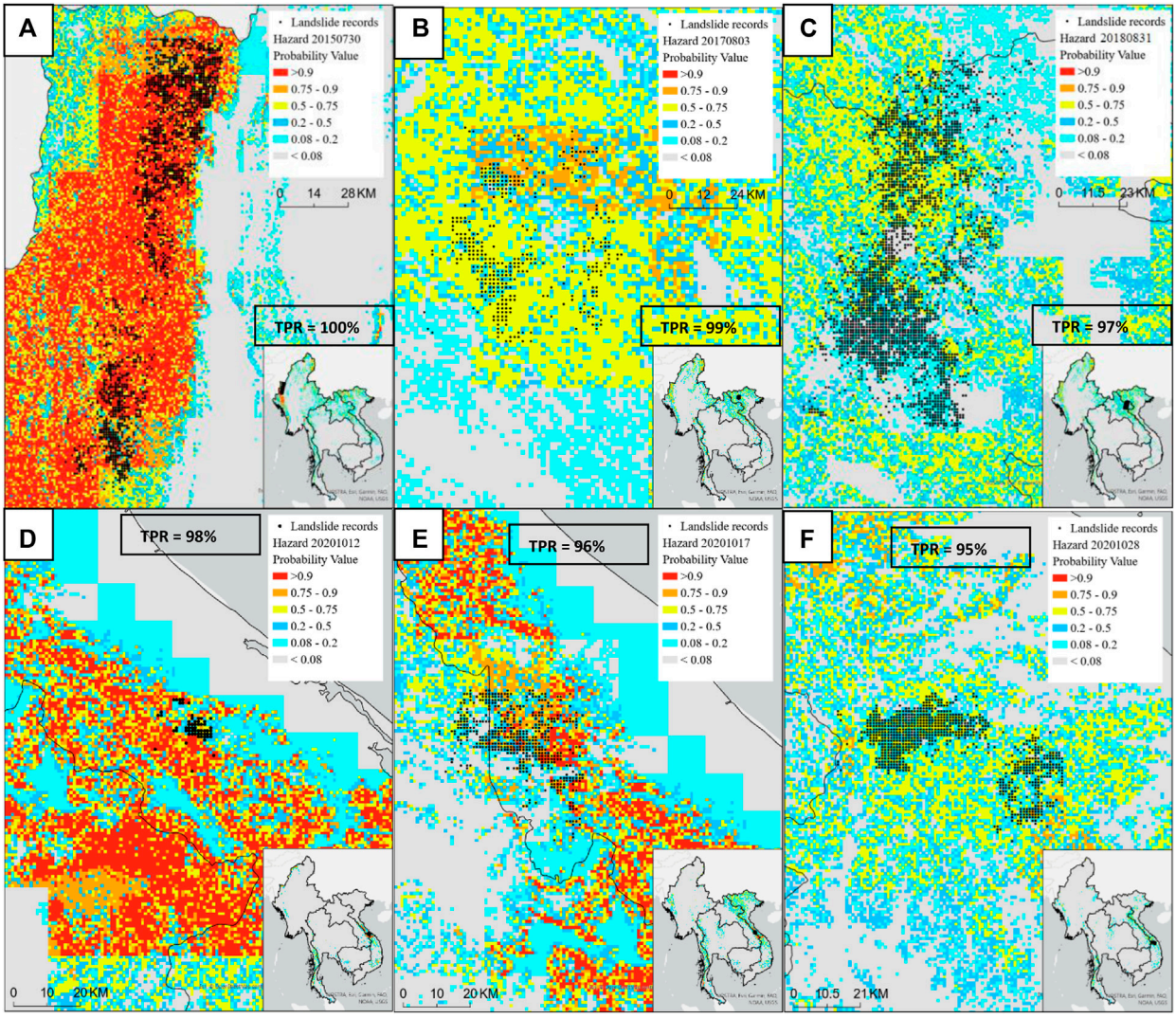


FIGURE 8
Hazard probability map with observed landslide records for different dates, **(A)** 30 July 2015 over Hakah-Falam, Myanmar; **(B)** 3 August 2017 over Muong La, Vietnam; **(C)** 31 August 2018 over Xieng Ngeun, Loas; **(D)** 12 October 2020 for the Phong Dien, Vietnam event, **(E)** 17 October 2020 over Huang Hua, Vietnam; and **(F)** 28 October 2020 for the Quang Nam, Vietnam event. The true positive rate (TPR) for each inventory is shown.

Model uncertainty also stems from variability and accuracy of the landslide triggering variables selected for the model. The precipitation triggering information from NASA GPM IMERG has known challenges in representing the total rainfall accumulations for short-duration, high intensity events as well as orographically-impacted events. We also consider CFS soil moisture product due to its near-real-time availability; however, the 25 km spatial resolution of the CFS product cannot adequately represent the actual soil moisture content at local spatial scales. Further, in this tropical environment validating this type of product is not feasible due to the dearth of *in situ* soil moisture information. Among the topographic variables, NASA DEM can be another source of uncertainty that could be propagated to slope and relief features of the model, with more recent topographic changes not accurately resolved in the data source on which the model was trained. Finally, the SALaD algorithm-generated inventory may not robustly cover all cover types where landslides may occur, which may bias the model validation. However, this is a challenge inherent in all data-limited environments and care was taken to ensure that we developed the most comprehensive inventory possible to support this work.

The false alarm rate generated by the model was higher than anticipated but is similar to the trend observed in the case of

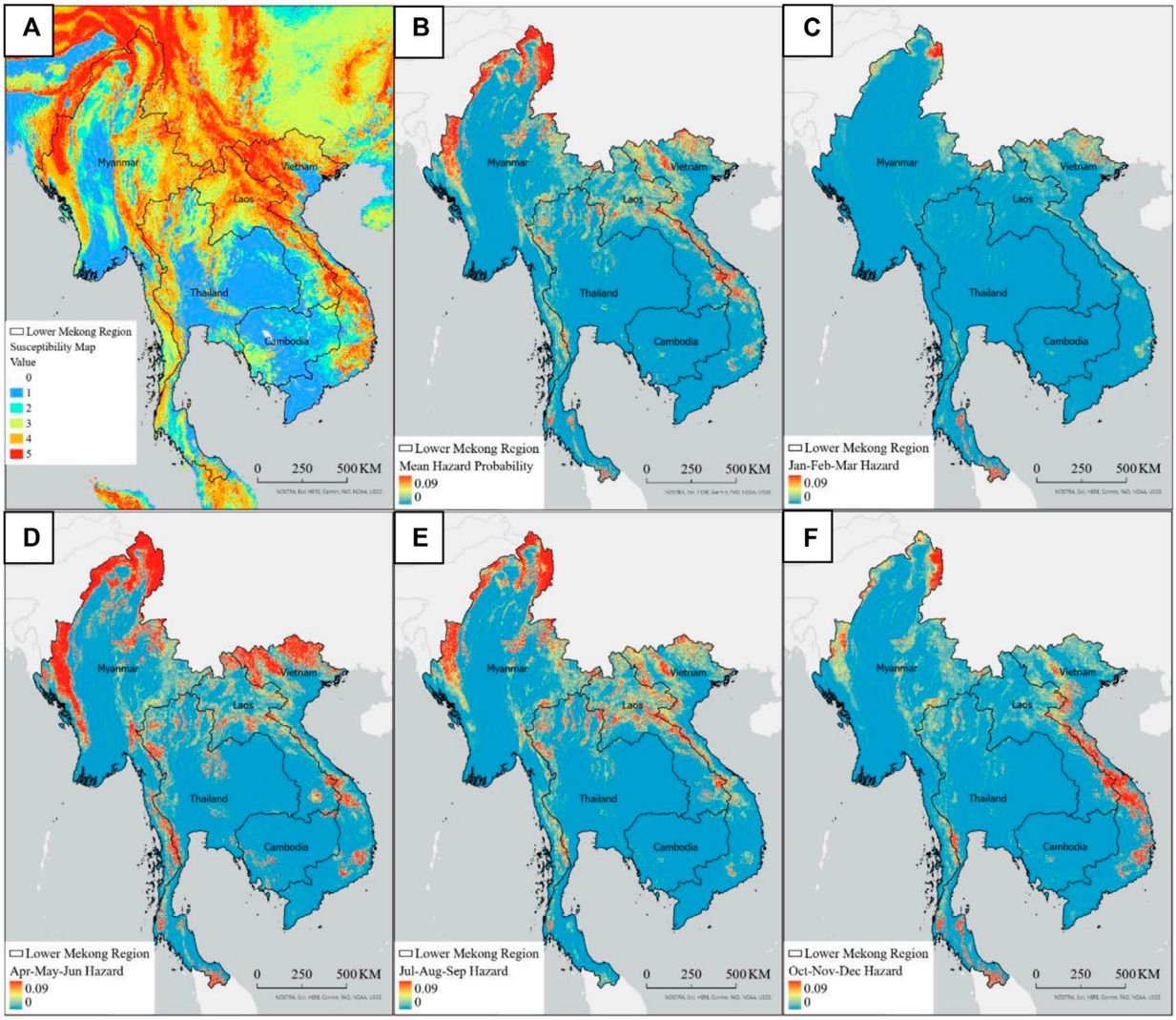


FIGURE 9
The geographic and seasonal distribution of landslide hazard across the LMR; **(A)** Global susceptibility map, **(B)** long-term mean hazard probability map, **(C)** Mean hazard probability map of Jan-March, **(D)** Mean hazard probability map of April-June, **(E)** Mean hazard probability map of July-September, **(F)** Mean hazard probability map of October-December.

Global LHASA model. The model solely depends on the triggering variables in estimating the landslide hazard probability. During the model training phase, random grids in the spatial-temporal direction were treated as non-landslide samples; however, the non-landslide areas cannot be corroborated in all cases. As such, the false alarm rate should be regarded as a relative estimate to represent the expansiveness of the potential hazard impact area for situational awareness rather than providing an accurate prediction of landslide events occurring at very fine resolution. Furthermore, the threshold defined in the model to distinguish between landslides and non-landslides played another role in a higher false positive rate for the tradeoffs. A well-established landslide inventory along with further testing on the threshold can minimize the false positive rate.

To reduce the dependency on the local computation resources, the modeling framework was implemented in the Google Cloud Platform. The uniqueness of this framework is that it was developed based on the most updated open-source cloud computing framework. The GCP provides a great opportunity in the terms of deployment, data and application programming interface (API) accessibility, reproducibility, and scalability. The model can be customized according to the available data, retrained and re-deployed to the same bucket in the easiest way just by using a web browser. An ongoing effort has focused on training and capacity building of the regional

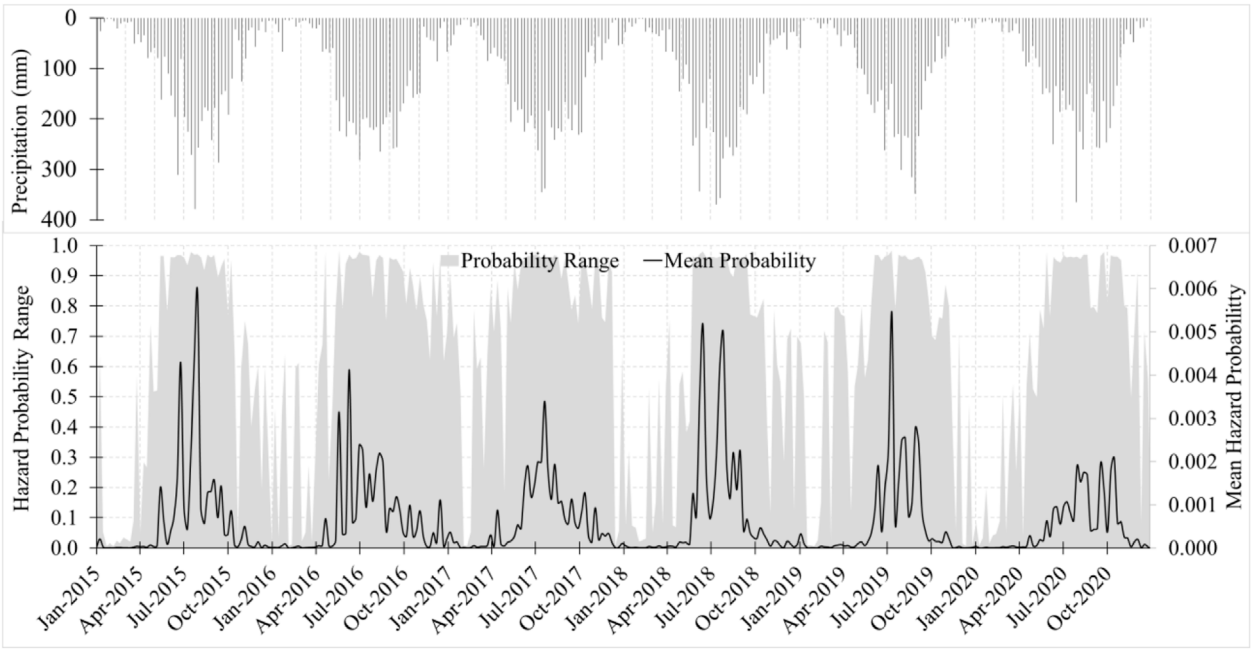


FIGURE 10
Long-term mean hazard probability timeseries at weekly scale with precipitation for the LMR.

agencies within the LMR to use LHASA-Mekong in order to harness the power of freely available cloud computing resources and the scalable framework for decision-making and emergency management.

4 Conclusion

Using globally available hydrologic and topographic variables and extensive new landslide inventories, LHASA-Mekong was trained, validated, revalidated, and deployed as part of this study. The model was designed to keep customizability in mind so that end-users can retrain and deploy it based on the accessibility to improved triggering variables and regional data. Currently, the model has been producing nowcast hazard probability, but work is ongoing to explore how forecast precipitation products such as Climate Hazards center InfraRed Precipitation with Station data - Global Ensemble Forecast System (CHIRPS-GEFS, Harrison et al., 2022). Another benefit of this open-source framework with publicly available source data is that it can be used for the other regions. However, care must be taken to ensure that data is sufficient, and the proper training and validation procedures are followed to ensure robust performance. Ultimately, this framework has tremendous potential to generate important landslide hazard assessments for data-sparse regions to inform emergency response, decision making, and planning to mitigate impacts within communities.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

DK, TS, PA, and PT contributed to the study conception and design. Material preparation, data collection and analysis were performed by NB, TS, and PA. The first draft of the manuscript was written by NB and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. There are no competing interests related to this work.

Funding

This research was funded by the NASA SERVIR Science Team (NNH18ZDA001N-18-SERVIR18_2-0036) and the collaboration with ADPC was supported *via* the joint

United States Agency for International Development (USAID) and NASA initiative SERVIR-Mekong.

Acknowledgments

We would also like to thank SERVIR-Mekong team at ADPC, SERVIR Coordination Office team, and Spatial Informatics Group for their valuable support throughout the study.

Conflict of interest

Author AP was employed by Spatial Informatics Group, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amatya, P., Kirschbaum, D., and Stanley, T. (2021b). Rainfall-induced landslide inventories for Lower Mekong based on Planet imagery and a semi-automatic mapping method. *Geoscience Data J.* doi:10.1002/gdj3.145

Amatya, P., Kirschbaum, D., Stanley, T., and Tanyas, H. (2021a). Landslide mapping using object-based image analysis and open source tools. *Eng. Geol.* 282, 106000. doi:10.1016/j.enggeo.2021.106000

Asian Disaster Preparedness Center (2008). Department of Mineral Resources, & Geotechnical Engineering Research and Development Center. Landslide Mitigation Demonstration Project for Patong City (Asian Program for Regional Capacity Enhancement for Landslide Impact Mitigation (RECLAIM II)) [Final Report]. Bankok, Thailand.

Biswas, N. K., and Hossain, F. (2018). A scalable open-source web-analytic framework to improve satellite-based operational water management in developing countries. *J. Hydroinformatics* 20 (1), 49–68. doi:10.2166/hydro.2017.073

Biswas, N. K., Hossain, F., Bonnema, M., Aminul, A., Biswas, R., Buiyan, A., et al. (2020). A computationally efficient flashflood early warning system for a mountainous and transboundary river basin in Bangladesh. *J. Hydroinformatics* 22 (6), 1672–1692. doi:10.2166/hydro.2020.202

Biswas, N. K., Hossain, F., Bonnema, M., Lee, H., and Chishtie, F. (2021). Towards a global Reservoir Assessment Tool for predicting hydrologic impacts and operating patterns of existing and planned reservoirs. *Environ. Model. Softw.* 140, 105043. doi:10.1016/j.envsoft.2021.105043

Bui, D. T., Lofman, O., Revhaug, I., and Dick, O. (2011). Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards (Dordr).* 59 (3), 1413–1444. doi:10.1007/s11069-011-9844-2

Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I., and Dick, O. B. (2012). Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput. Geosciences* 45, 199–211. doi:10.1016/j.cageo.2011.10.031

Bui, D. T., Tuan, T. A., Hoang, N. D., Thanh, N. Q., Nguyen, D. B., Van Liem, N., et al. (2017). Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* 14 (2), 447–458. doi:10.1007/s10346-016-0711-9

Dang, K., Loi, D. H., Sassa, K., Duc, D. M., and Ha, N. D. (2020). "Hazard assessment of a rainfall-induced deep-seated landslide in Hakha city, Myanmar," in *Workshop on world landslide forum* (Cham: Springer), 249–257. doi:10.1007/978-3-030-60706-7-23

GEM Hazard Team (2019). GEM global active faults. Available at: https://github.com/GEMScienceTools/gem-global-active-faults/tree/master/geopackage (Accessed July 19, 2019).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi:10.1016/j.rse.2017.06.031

Harrison, L., Landsfeld, M., Husak, G., Davenport, F., Shukla, S., Turner, W., et al. (2022). Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. *Sci. Data* 9, 375–413. doi:10.15780/G2PH2M

Hartmann, J., and Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the earth surface. *Geochem. Geophys. Geosyst.* 13. doi:10.1029/2012GC004370

Hearn, G. J., and Pongpanya, P. (2021). Developing a landslide vulnerability assessment for the national road network in Laos. *Q. J. Eng. Geol. Hydrogeo.* 54 (3). doi:10.1144/qjegh2020-110

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., et al. (2015). NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) version. 4 (26).

Intarawichian, N., and Dasananda, S. (2011). Frequency ratio model based landslide susceptibility mapping in lower Mae Chaem watershed, Northern Thailand. *Environ. Earth Sci.* 64 (8), 2271–2285. doi:10.1007/s12665-011-1055-3

Jotisankasa, A., and Vathananukij, H. (2008). "Investigation of soil moisture characteristics of landslide-prone slopes in Thailand," in International conference on management of landslide hazard in the Asia-Pacific region, November 11–15, 2008, 12

Kanjanakul, C., Chub-uppakarn, T., and Chalermyanont, T. (2016). Rainfall thresholds for landslide early warning system in Nakhon Si Thammarat. *Arab. J. Geosci.* 9 (11), 584. doi:10.1007/s12517-016-2614-4

Kirschbaum, D., Stanley, T., and Zhou, Y. (2015). Spatial and temporal analysis of a global landslide catalog. *Geomorphol.* 249, 4–15. doi:10.1016/j.geomorph.2015.03.016

Kolomazník, J., Hlavacova, I., and Schloegl, M. (2021). Supporting disaster risk reduction with satellite Earth Observation Landslide hazard assessment for the Chin road corridor, Myanmar. In EGU General Assembly Conference Abstracts (pp. EGU21-1476). doi:10.5194/egusphere-egu21-14769

Komori, D., Rangsiwanichpong, P., Inoue, N., Ono, K., Watanabe, S., and Kazama, S. (2018). Distributed probability of slope failure in Thailand under climate change. *Clim. Risk Manag.* 20, 126–137. doi:10.1016/j.crm.2018.03.002

Lee, J. K., Kim, H. S., Kuk, S. H., and Park, S. W. (2006). in *Development of an e-Engineering environment based on service-oriented architectures. Cooperative Design, Visualization, and Engineering.* Editor Y. Luo. LNCS 4101. doi:10.1007/s00254-006-0256-7

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., and Schipper, A. M. (2018). Global patterns of current and future road infrastructure. *Environ. Res. Lett.* 13, 064006. doi:10.1088/1748-9326/aabd42

Naing, M. T., Aye, M. M., and Oo, K. L. (2022). Rainfall and landslide susceptibility in Hakha environ in northern chin state, Myanmar. *Br. J. Arts Humanit.* 4 (1), 01–14. doi:10.34104/bjah.02201014

Nawagamuwa, U. P., Bhasin, R. K., Kjekstad, O., and Arambepola, N. M. S. I. (2013). "Recommending regional rainfall threshold values for early warning of landslides in the asian region," in *Landslide science and practice: Volume 4: Global environmental change*. Editors C. Margottini, P. Canuti, and K. Sassa (Springer), 235–241. doi:10.1007/978-3-642-31337-0-30

Oh, H. J., Lee, S., Chotikasathien, W., Kim, C. H., and Kwon, J. H. (2009). Predictive landslide susceptibility mapping using spatial information in the Pechabun area of Thailand. *Environ. Geol.* 57 (3), 641–651. doi:10.1007/s00254-008-1342-9

Ohtsu, H., Chaleiwchalard, N., Koga, H., and Soralump, S. (2012). "A study on landslide early warning system considering the effect of antecedent rainfall on slope stability," in *The 41st symposium on rock mechanics*, 252–257.

Ono, K., Kazama, S., and Ekkawatpanit, C. (2014). Assessment of rainfall-induced shallow landslides in Phetchabun and Krabi provinces, Thailand. *Nat. Hazards (Dordr).* 74 (3), 2089–2107. doi:10.1007/s11069-014-1292-3

Panday, S., and Dong, J. J. (2021). Topographical features of rainfall-triggered landslides in Mon state, Myanmar, August 2019: Spatial distribution heterogeneity and uncommon large relative heights. *Landslides* 18 (12), 3875–3889. doi:10.1007/s10346-021-01758-7

Phien-Wej, N., Nutalaya, P., Aung, Z., and Zhibin, T. (1993). Catastrophic landslides and debris flows in Thailand. *Bull. Int. Assoc. Eng. Geol.* 48 (1), 93–100. doi:10.1007/bf02594981

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., and Becker, E. (2011). NCEP climate forecast system version 2 (CFSv2) 6-hourly products. *Res. Data Archive A. T. Natl. Cent. Atmos. Res. Comput. Inf. Syst. Laboratory* 10. doi:10.5065/D61C1TXF

Salee, R., Chinkulkijniwat, A., Yubonchit, S., Horpibulsuk, S., Wangfaoklang, C., and Soisompong, S. (2022). New threshold for landslide warning in the southern part of Thailand integrates cumulative rainfall with event rainfall depth-duration. *Nat. Hazards (Dordr).* 113, 125–141. doi:10.1007/s11069-022-05292-0

Schmidt-Thomé, P., Tatong, T., Kunthasap, P., and Wathanaprida, S. (2018). Community based landslide risk mitigation in Thailand. *Episodes* 41 (4), 225–233. doi:10.18814/epiiugs/2018/018017

Sidle, R. C., and Ziegler, A. D. (2017). The canopy interception–landslide initiation conundrum: Insight from a tropical secondary forest in northern Thailand. *Hydrol. Earth Syst. Sci.* 21 (1), 651–667. doi:10.5194/hess-21-651-2017

Soralump, S. (2010). Geotechnical approach for the warning of rainfall-triggered landslide in Thailand considering antecedence rainfall data. International Conference on Slope 2010: Geotechnique and Geosynthetics for Slopes, Chiangmai, Thailand

Stanley, T., and Kirschbaum, D. B. (2017). A heuristic approach to global landslide susceptibility mapping. *Nat. Hazards* 87 (1), 145–164. doi:10.1007/s11069-017-2757-y

Stanley, T. A., Kirschbaum, D. B., Benz, G., Emberson, R. A., Amatya, P. M., Medwedeff, W., et al. (2021). Data-driven landslide nowcasting at the global scale. *Front. Earth Sci. (Lausanne).* 378. doi:10.3389/feart.2021.640043

Titti, G., Borgatti, L., Zou, Q., Cui, P., and Pasuto, A. (2021). Landslide susceptibility in the belt and road countries: Continental step of a multi-scale approach. *Environ. Earth Sci.* 80, 630. doi:10.1007/s12665-021-09910-1

Tran, Q. C., Minh, D. D., Jaafari, A., Al-Ansari, N., Minh, D. D., Van, D. T., et al. (2020). Novel ensemble landslide predictive models based on the hyperpipes algorithm: A case study in the nam dam commune, vietnam. *Appl. Sci.* 10 (11), 3710. doi:10.3390/app10113710

Truong, X. L., Mitamura, M., Kono, Y., Raghavan, V., Yonezawa, G., Truong, X. Q., et al. (2018). Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl. Sci.* 8 (7), 1046. doi:10.3390/app8071046

Van Dao, D., Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., et al. (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* 188, 104451. doi:10.1016/j.catena.2019.104451

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/feart.2022.1009617

OPEN ACCESS

EDITED BY
Zefa Yang,
Central South University, China

REVIEWED BY
Xiaojun Zhu,
Anhui University, China
Wang Lei,
Anhui University of Science and Technology, China

*CORRESPONDENCE
Xinguo Zhang,
zhangxg1229@163.com

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 02 August 2022
ACCEPTED 22 September 2022
PUBLISHED 09 January 2023

CITATION
Meng Z, Dong Y, Zhang X, Jiao F, Fan Y and Thammavongsa C (2023), Short-wall paste continuous mining and continuous backfilling for controlling industrial square surface subsidence. *Front. Earth Sci.* 10:1009617. doi: 10.3389/feart.2022.1009617

Short-wall paste continuous mining and continuous backfilling for controlling industrial square surface subsidence

Zixuan Meng[1], Yue Dong[2], Xinguo Zhang[1]*, Fangshu Jiao[2], Yaqi Fan[2] and Champa Thammavongsa[2]

[1]College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao, China, [2]College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao, China

To facilitate coal mining under the industrial square in Western China, short-wall paste continuous mining and continuous backfilling (PCMCB) is proposed to release the overlaid resources and handle coal gangue. First, the concept of short-wall PCMCB is presented, and the principle of surface subsidence is analyzed. Second, the mechanical model of the branch roadway and strength of the backfill body are established via mechanical derivations. Third, the proportion of backfill materials is calculated. Finally, field applications are carried out in the Dingjiaqu coal mine. The results show the following: 1) Based on the simply supported beam model of the roof and strength design model of the backfill body, the reasonable width of the branch roadway was determined as 5 m and late strength of the backfill body should exceed 2.62 MPa in the Dingjiaqu coal mine. 2) According to the design strength of the backfill body and backfill material proportion, the reasonable ratio of cement: fly ash: gangue: water was determined as 1:1.82:4.17:1.78. 3) In accordance with the equivalent mining height theory, the critical filling rate of the goaf should be more than 90% to ensure that the deformation is within the first level of building damage. 4) The field application shows that subsidence could be effectively controlled by backfilling, the overall surface subsidence was very small, and the maximum subsidence value (77 mm) was far lower than the deformation extremes of the building protection requirements. This study provides practical guidance for the exploitation of overlaid resources under similar shallow burial conditions, which can help in the disposal of coal gangue with considerable social and economic benefits with popularization and application.

KEYWORDS

paste backfill, surface subsidence, solid waste utilization, shallow coal seam, continuous mining and continuous backfilling

1 Introduction

China has abundant coal resources that are widely distributed across the country. The overall distribution of the coal resources is uneven, with more availability in the west and less in the east. Coal resources in western China account for more than 70% of total national coal resources; with the gradual depletion of the coal resources in eastern China, the west is expected to be the main production base for coal and is characterized by its large scale, high intensity, high efficiency, and high recovery rate (Han et al., 2018; Xiao et al., 2021). Given the arid or semi-arid climate and inadequate water resources, the ecological environment in the west is fragile and coal seams are generally shallow (Yang et al., 2018). Traditional fully mechanized caving mining causes greater damage to the surface and environment, and the low recovery rate of partial mining does not conform with the concept of sustainable development (Fan and Ma, 2018; Yang et al., 2019). In parts of the west, the amounts of "three under" coal resources (i.e., those under buildings, railroads, and waterways) are large, and the problem of coal overlaid under industrial squares is typical. The amount of coal overlaid under the industrial square is estimated to be of the order of millions to tens of millions of tons, but there are many buildings in the industrial square with complex structures and high protection levels (Feng et al., 2020). Therefore, solving the problem of coal overlaid under the industrial square is important for solving the problem of three under mining in western China; this will also entail milestone significance for achieving large-scale resource recovery (Yan et al., 2018). Backfill mining is a technology developed in the past decade and can be used to control surface subsidence as well as protect buildings and structures in the industrial square by reducing the equivalent mining height. At the same time, the backfill method with coal gangue as the main raw material can effectively solve the problem of gangue disposal while protecting the fragile ecological environment in west China (Zhang et al., 2021a; Li et al., 2021).

The mainstream backfill technologies used in China are solid, paste, high water, and slurry backfill technologies. The application conditions of each of these backfill technologies are different. The purpose of backfilling is fundamental to the choice of the backfill technology used. If backfilling is adopted for solving problems with the three under coal resources, the primary purpose is to control subsidence; if it is used to solve problems with gangue disposal, then the main purpose is to digest solid wastes. At the same time, particular mining conditions must be considered in the backfill process because the proportion of backfill materials required in each mine as well as the model used are different.

Zhang Q. et al. (2019) summarized the research progress on solid backfill technologies in China and introduced the mining-separating-backfilling technology based on the specific engineering case of the Pingmei no.12 coal mine. Zhu et al. (2016) proposed a model to accurately predict the surface dynamic subsidence process of solid backfill mining and assess the mining damage; the validity of this method was verified using in situ measured subsidence data from the roofs of two backfill working faces. Zhang and Wang (2007) studied the feasibility of cemented gangue backfill and noted that coal gangue was not an ideal aggregate for pipeline gravity flow backfill; however, the disadvantages of gangue, such as bad fluidity and severe pipe wear, could be overcome by the addition of fly ash to obtain reasonable proportions of the backfill materials. Belem and Benzaazoua (2007) presented the laboratory- and backfill-plant-scale designs for paste backfill mix design and optimization, and they built different models (analytical, semi-empirical, and empirical) to predict the rheological factors of paste backfill (shear yield stress and viscosity). Chang et al. (2021) created a mechanical model of a paste false roof to analyze its stability for lower slicing mining and obtained the law of the plastic zone in the lower slicing face. Based on the characteristics of strip mining and backfilling, Zhu et al. (2007) proposed the method of filling with grouting for mining-induced overburden separation, which was successfully used without village relocation in the Huaibei mining area. Based on the properties of ultrahigh water materials, Jia and Feng (2012) analyzed the basic features of high water materials and presented different high water backfill methods for a goaf as engineering cases. By employing aeolian sand as the backfill aggregate, Bai et al. (2018) studied the mechanical properties of paste backfill materials with different proportions to provide reference for backfilling of the shallow seam in western China.

Extensive studies have been carried out to study the feasibility of using solid wastes, such as coal gangue, to recycle the overlaid coal resources owing to the increasing need for energy and consciousness regarding environmental protection. Great contributions on the backfill models and materials have been made in the development of backfill mining technologies by the above scholars, and good application results have been achieved in the field with various backfill technologies, especially in longwall mining. However, the filling rate of solid backfill is generally less than 80%, and its control of surface subsidence is not as good as those of paste and high water backfill; solid backfill is mainly used for digesting quantities of gangue. Although high water backfill can fulfill the control requirements of surface deformation, it cannot be used to dispose large amounts of solid wastes; spaces for separated strata need to be formed in slurry backfill, but the current research on delaminated space development are insufficient (Zhang J. et al., 2019). In response to the above problems, paste backfill is proposed as a reasonable means to achieve both high filling rate and effective digestion of gangue. The application of backfill mining is limited by restrictions on various factors, such as filling cost and filling yield. According to statistics, backfill mines are mainly distributed in eastern China in the Shandong, Shanxi, and Hebei provinces. There are very few cases of coal recovery

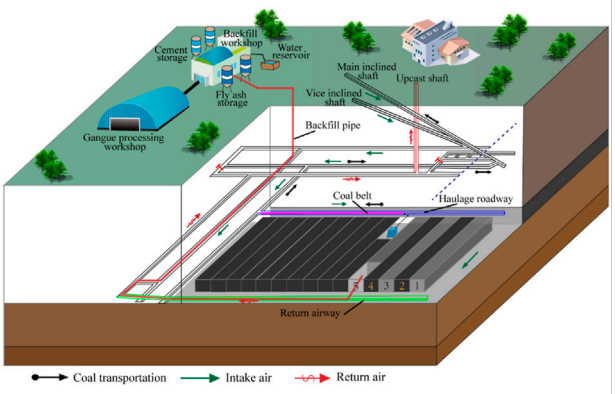


FIGURE 1
Diagram of the short-wall PCMCB system.

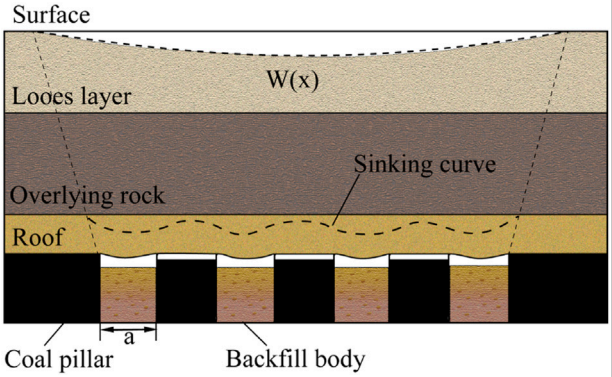


FIGURE 2
Surface subsidence model of short-wall PCMCB.

under the industrial squares in western China (Liu et al., 2020). This study is mainly concerned with the short-wall PCMCB technology for developing resources under the industrial square. The basic concept is first presented, and the principle of controlling surface subsidence is analyzed; the mechanical model of the branch roadway and strength of the backfill body are established next; based on the geological mining conditions of the Dingjiaqu coal mine, the parameters of the branch roadway and strength of the backfill body are determined, and the reasonable proportions of backfill materials are calculated; finally, the critical filling rate for controlling deformation of the building structure is obtained from the equivalent mining height theory, and the feasibility of this technology is proved through a site application.

2 Concept and principle

2.1 Concept of short-wall PCMCB

Short-wall PCMCB is a green mining technology that is being used and developed since 2012. The diagram of the short-wall PCMCB system is shown in Figure 1, and the system is mainly divided into the gangue crushing and storage, stirring, pipeline pumping, and coal mining and separating systems. Coal gangue is transported to the backfill workshop after crushing; then, it is placed in a blender with fly ash, cement, and water to form a paste according to the determined proportion. Finally, the paste is transported to the goaf through pipelines. Before the backfill process, a complete ventilation and transportation system should be implemented on the basis of the original development and mining roadway. Then, the mining area is divided into multiple branch roadways in a direction parallel to the open-off cut and numbered serially as 1, 2, 3, 4, 5, etc. The odd-numbered branch roadways are mined first; when the filled odd-numbered branch roadways are stable, the remaining even-numbered branch roadways are mined so as to complete the mining and backfilling process (Lin et al., 2021). Short-wall PCMCB technology effectively integrates the three machines matching of short-wall mining and paste backfill process; it realizes separation of mining and backfilling while avoiding technical problems, such as mutual interference, low efficiency, and poor roof control in longwall mining and backfilling (Zhang et al., 2021b).

2.2 Principle of controlling surface subsidence by short-wall PCMCB

According to the basic principle of mining subsidence, when the overburden lithology is certain, a smaller mining width a leads to a lower mining degree, and a smaller surface movement and deformation are achieved (Kang et al., 2019); the surface subsidence model is shown in Figure 2. Under the action of the overburden load, a combined support is formed by the backfill body and coal pillar, which hinders further deformation of the goaf and maintains the surface stable. Through PCMCB, mining and backfilling can be carried out simultaneously; thus, the area is always in a state of subcritical mining so as to prevent large subsidence deformations on the surface. This is the principle of short-wall PCMCB for controlling surface subsidence.

3 Design of technical parameters

3.1 Design of the width and length of the branch roadway

In the width design, the immediate roof can be regarded as a beam structure supported by coal pillars on both sides. Because the buried

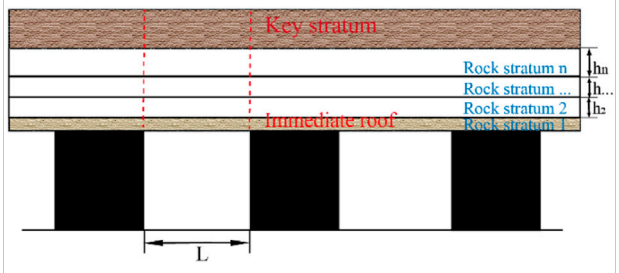


FIGURE 3
Model of the composite beams.

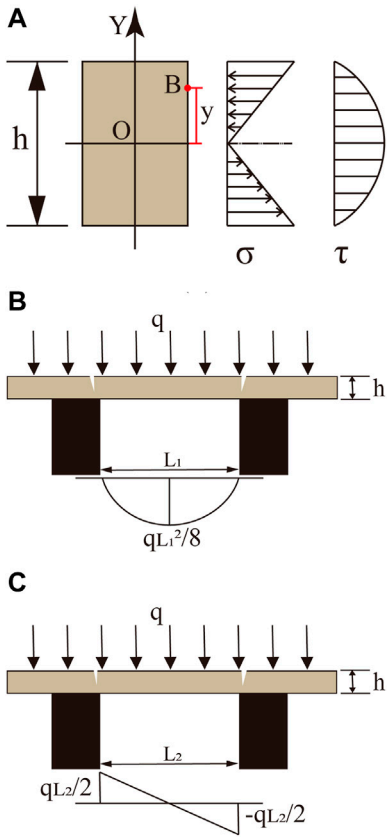


FIGURE 4
Failure model of a simply supported beam under a uniform load for **(A)** the stress state of any point on the section, **(B)** bending failure, and **(C)** shear failure.

depth of the coal seam is low and the coal pillar is affected by mining, the stability calculation of the roof can be considered as a simply supported beam (Chi, 2012). The calculation of the roof load q is the premise of rock beam stability calculation; according to the principle of a composite beam (Figure 3), this can be expressed as (Qian et al., 1996)

$$(q_n)_1 = \frac{E_1 h_1^3 (\gamma_1 h_1 + \gamma_2 h_2 + \cdots + \gamma_n h_n)}{E_1 h_1^3 + E_2 h_2^3 + \cdots + E_n h_n^3} \quad (1)$$

In the above formula, $(q_n)_1$ is the load imposed by the nth layer stratum controlled by the immediate roof [Pa]; E_n is the elastic modulus of the nth layer rock stratum [Pa]; γ_n is the volumetric weight of the nth layer rock stratum [N/m^3]; h_n is the thickness of the nth layer rock stratum [m].

When $(q_n)_1 < (q_{n-1})_1$ is calculated, $(q_n)_1$ is used as the load applied to the first layer. As shown in Table 5, $(q_1)_1 = 0.1264$ MPa, $(q_2)_1 = 0.0178$ MPa, and roof load $q = 126.4$ kPa.

In accordance with the stress distribution of a simply supported beam under a uniform load, the normal and shear stresses at any point B (Figure 4A) can be expressed as (Gere and Goodno, 2012)

$$\sigma_x = \frac{12 M_x y}{h^3}, \quad (2)$$

$$\tau_{xy} = \frac{3 Q_x (h^2 - 4y^2)}{h^3}, \quad (3)$$

where M_x and Q_x are the bending moment and shear force in the cross section on which point B is located; y is the distance from point B to the neutral axis; h is the height of the beam.

Taking into account the nonhomogeneity of rock, a safety factor N is considered for the ultimate strength of the rock beam, and the allowable normal stress $[\sigma]$ and shear stresses $[\tau]$ can be expressed as

$$[\sigma] = \frac{R_t}{N}, \quad (4)$$

$$[\tau] = \frac{R_j}{N}, \quad (5)$$

where N is the safety factor that ranges from 2 to 4; R_t is the tensile strength [MPa]; R_j is the shear strength [MPa].

When bending damage of a rock beam occurs, the maximum bending moment is at the center of the beam (Figure 4B), and the maximum normal stress can be expressed as (Gere and Goodno, 2012)

$$\sigma_{max} = \frac{3 q L_1^2}{h^2}. \quad (6)$$

In this case, the ultimate span of the rock beam can be expressed as

$$L_1 = h \sqrt{\frac{4[\sigma]}{3q}}. \quad (7)$$

When shear damage of a rock beam occurs, the maximum shear force appears at both ends of the beam (Figure 4C), and the maximum shear stress generated can be expressed as (Gere and Goodno, 2012)

$$\tau_{max} = \frac{3 q L_2}{4h}. \quad (8)$$

TABLE 1 Parameters of the rock strata and surface movement of the 42201 working face.

Subsidence coefficient/η	Tangent of major influence angle/tanβ	Horizontal displacement factor/b	Deviation of inflection point/s	Propagation angle of extraction/θ_0
0.65	1.77	0.32	0.2	88°

TABLE 2 Surface deformation characteristics affected by the filling rate.

Filling rate (%)	Surface subsidence (mm)	Inclined deformation (mm/m)		Horizontal deformation (mm/m)		Curvature (mm/m^2)	
		E-W direction	S-N direction	E-W direction	S-N direction	E-W direction	S-N direction
0	1,669	38.14/	49.05/	13.68/	20.32/	1.30/	1.92/
		−15.51	−55.28	−16.44	−24.08	−1.34	−2.34
60	668	15.26/	19.62/	5.47/	8.13/	0.52/	0.77/
		−6.21	−22.11	−6.57	−9.63	−0.53	−0.93
80	334	7.63/	9.81/	2.73/	4.06/	0.26/	0.38/
		−3.10	−11.06	−3.29	−4.81	−0.27	−0.47
90	167	3.81/	4.9/	1.37/	2.03/	0.13/	0.19/
		−1.55	−5.53	−1.64	−2.41	−0.13	−0.23
95	83	1.91/	2.45/	0.68/	1.02/	0.07/	0.10/
		−0.78	−2.76	−0.82	−1.20	−0.07	−0.12

Owing to shear damage, the ultimate span of the rock beam can be expressed as

$$L_2 = \frac{4h[\tau]}{3q}. \tag{9}$$

Referring to Table 5, q, R_t, and R_j were substituted into the above Eqs 7, 9 to obtain L_1 = 13.07 m andL_2 = 49.14 m. Considering the site process design and other factors, the final branch roadway and coal pillar width are 5 m each.

The length of the branch roadway can be solved using the lowest cost per ton of coal, which can be expressed as (Liu et al., 2007)

$$Z = Z_1 + Z_2 + Z_3 = \frac{a}{l} + \frac{(b+c)}{l+\Delta}, \tag{10}$$

where Z_1, Z_2, and Z_3 are the roadway tunneling, transportation, and maintenance costs, respectively; a, b, and c are the tunneling, transportation, and maintenance cost factors, respectively; l represents the length of the branch roadway; Δ represents the ancillary cost.

From the formula, it is seen that the longer the branch roadway, the lower is the tunneling cost. Correspondingly, the costs of transportation and maintenance of the roadway will also increase. The length of the branch roadway also depends on the mining equipment; in the Dingjiaqu coal mine, two EBZ-160 fully mechanized driving machines with 40T scraper conveyors are employed, and the maximum daily excavation capacity is 40 m. To reduce the amount of continuous handling of the equipment, the final length of the branch roadway is 200 m.

3.2 Strength design of the backfill body

The strength of the backfill body generally comprises early and late strengths. There is no concept of early strength for the short-wall PCMCB as the backfill materials are fully solidified in the odd-numbered branch roadways when the even-numbered branch roadways are to be mined. The maximum strength of the backfill body is required to recover the coal pillar safely under its supporting condition. Therefore, the late strength design is the key to the backfill body. At present, the effective area theory is mainly used to analyze the load sustained by the backfill body; the backfill body bears the gravity of the rock layers directly above it and half of the branch roadway on each side. The load P borne by the backfill body is calculated by the following formula (Xie et al., 1998):

$$P = (a+b)\gamma\frac{H}{a}, \tag{11}$$

where P is the average stress borne by the coal pillar [MPa]; γ is the average weight of the overlying seam [MN/m^3]; H is the

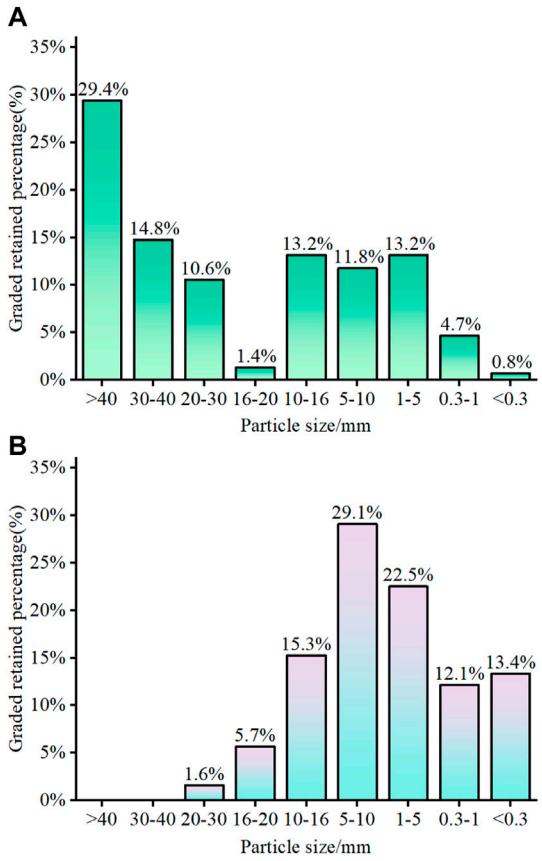


FIGURE 5
Different particle size percentages of gangue: **(A)** washed gangue; **(B)** crushed gangue.

average mining depth [m]; a and b are the widths of the reserved and mining coal pillars, respectively.

The late strength is usually calculated by the Bieniawski formula (Bieniawski and Van Heerden, 1975), which is expressed as

$$\sigma_p = \sigma_m\left(0.64 + 0.36\frac{W}{h}\right)^n, \tag{12}$$

where σ_p is the permissible strength of the backfill body; σ_m is the uniaxial strength of the critical cube specimen; W is the width of the backfill body; h is the height of the backfill body; n is a constant that depends on the ratio of width to height of the coal pillar; when $W/h>5$, $n=1.4$, and when $W/h<5$, $n=1$.

Considering the influence of the specimen size (Shao et al., 2018), σ_m can be expressed as

$$\sigma_m = \sigma_c\sqrt{\frac{D}{L}}, \tag{13}$$

where σ_c is the uniaxial compressive strength of the coal sample in the laboratory; D is the diameter or cube edge length of the laboratory coal sample; L is the limit size of the laboratory backfill sample.

Based on the limit strength theory, to maintain the stability of the backfill body, it is indispensable to consider the safety factor F such that $F = \frac{\sigma_p}{P} > 1.5–2.0$. The following equation can be deduced by combining Eqs 11–13.

$$\sigma_c \geq \frac{F\gamma H \cdot \frac{(a+b)}{a}}{\sqrt{\frac{D}{L}}\cdot\left(0.64 + 0.36\frac{W}{h}\right)^{1.4}}. \tag{14}$$

According to the actual state, $\gamma = 0.025\,\mathrm{MN/m^3}$, $H = 60\,m$, $a = b = 15\,m$, and $P = 3\,\mathrm{MPa}$; for $D = 50\,mm$, $L = 160\,mm$, and $F = 2$, we get $\sigma_c \geq 2.62\,\mathrm{MPa}$.

3.3 Equivalent mining height prediction

To determine whether the surface deformation after backfilling meets the protection requirements of buildings, it is necessary to predict the surface deformation accurately. The probability integral method is the most widely used approach to subsidence prediction (Wang et al., 2018). Under the gravity of the overburden strata, a certain deformation is produced between the coal roof and floor, and the backfill body will dehydrate and

TABLE 3 Uniaxial compression test results for different ratios of the backfill body over the ground.

No.	Curing period/d	Cement/kg	Fly ash/kg	Coal gangue/kg	Water/kg	Compressive strength/MPa
GT1	3	14	34	52	26	2.3
GT1	28	14	34	52	26	6.2
GT1	28	14	34	52	26	6.8
GT1	28	14.3	26	59.6	25.4	6.7
GT3	28	14.6	12.7	72	25.89	2.4
GT3	28	14.6	12.7	72.6	28.2	3.2
GT3	7	14.6	12.7	72.66	29	2.1
GT4	15	14.6	12.7	72.6	25	1.6
GT5	28	14.6	12.7	72.6	28.29	1.5

TABLE 4 Uniaxial compression test results for different ratios of the backfill body under the ground.

No.	Curing period/d	Cement/kg	Fly ash/kg	Coal gangue/kg	Water/kg	Compressive strength/MPa
GT2	65	14.62	12.77	72.6	28	4.2
GT2	65	14.62	12.77	72.6	25	2.98
GT4	55	14.6	26	59.5	25.4	3.13
GT6	50	14.6	26	59.5	25	5.08
GT8	42	14.62	12.77	72.6	28	4.2
GT12	23	14.6	12.7	72.6	24	2.86

TABLE 5 Rock mechanics parameters of the Dingjiaqu coal mine.

Lithology	Thickness (m)	Depth (m)	Capacity (MN/m^3)	E (GPa)	R_t (MPa)	R_j (MPa)
Surface soil	2.7	2.7	0.012	-	-	-
Loess	9.5	12.2	0.0125	-	-	-
Siltstones	3.14	15.34	0.0249	19.5	1.84	2.72
Sandy mudstone	6.82	22.16	0.0255	9.4	0.79	1.34
Siltstone	5.2	27.36	0.0247	19.5	1.84	2.72
Fine sandstone	8.04	35.4	0.0284	28.5	1.29	2.14
Siltstone	11.16	46.56	0.0247	19.5	1.84	2.72
Sandy mudstone	14.84	61.4	0.0253	9.4	0.79	1.34
Siltstone	5.14	66.54	0.0246	19.5	1.84	2.72
Coal seam	2.57	69.11	0.0125	8.3	0.45	0.77
Sandy mudstone	6.5	75.61	0.0253	9.4	0.82	1.36
Siltstone	1.05	76.66	0.0245	19.5	1.96	3.12
Coarse sandstone	8	84.66	0.023	11.2	1.21	1.86

TABLE 6 Key parameters of the branch roadway and backfill body in the Dingjiaqu coal mine.

Width/m	Length/m	Late strength/MPa	Ratio (cement: fly ash: gangue: water)	Critical filling rate/%
5	200	≥2.66	1:1.82:4.17:1.78	>90

shrink during the backfill process. Therefore, in actual calculations (Qu et al., 2010), the equivalent mining height theory is generally used to predict surface deformation, and the equivalent mining height is expressed as follows:

$$m_d = \delta + \Delta + s, \tag{15}$$

where δ is the movement quantity of the roof and floor, Δ is the deficit distance of roof contact for the backfill body, and s is the deformation of the backfill body.

Based on the industrial square building protection grade and equivalent mining height theory, the surface movement and deformation parameters of the 42201 working face (Table 1) were selected, and the surface deformation characteristics under different filling rates (equivalent mining height divided by actual mining height) were obtained using the probability integral method (Table 2). The results show that when the filling rate is more than 90%, the damage may be within the first level of building damage.

4 Backfill materials

Paste backfill materials are mainly composed of gangue, cement, fly ash, and water. As the most abundant material, gangue plays the role of the skeleton in the backfill body; as

FIGURE 6
Diagram showing the physical effects of a backfill plant.

FIGURE 7
Design and construction of a paste baffle.

the cementitious material, cement is the fundamental source of strength; fly ash can significantly reduce the amount of cement required and effectively improve the pumping performances of the backfill materials.

4.1 Reasonable gradation of gangue

Washed gangue from the Dingjiaqu coal washery was used in the field. After screening, the size distribution of the washed gangue is as shown in Figure 5A. Since the particle sizes of the washed gangue did not meet the requirements of long-distance pipeline transportation, the washed gangue had to be crushed until the maximum particle size was less than 25 mm. The particle size distribution after screening for the crushed gangue is as shown in Figure 5B. It can be seen that most of the crushed gangue particles are between 0.3 mm and 20 mm, and only a very small fraction of the particles are 20–25 mm in size. The gangue particles between 5 and 20 mm account for 50.1%, and those with sizes less than 5 mm account for 48%, which conforms to the ideal particle gradation curve (maximum particle size of the gangue is less than 25 mm, and the proportion of gangue with particle sizes less than 5 mm is about 40–50%) (Zheng et al., 2006; Sun et al., 2017).

4.2 Design of the proportion of backfill materials

Since there are many studies on the proportion of backfill materials (Koohestani et al., 2016; Jiang et al., 2020; Liu et al., 2022), the composition was adjusted according to the one in the neighboring Tianyu coal mine, and the laboratory and downhole proportioning experiments were conducted. As shown in Tables 3, 4, the final results obtained via the uniaxial compression test show that when the cement dosage is certain, excess gangue leads to lower strength of the backfill body, so the proportion of gangue should be in a reasonable range. The final choice of backfill material proportion for cement: fly ash: gangue: water is 1:1.82:4.17:1.78 for the Dingjiaqu coal mine.

5 Case studies

5.1 Geological settings and mining conditions

The Dingjiaqu coal mine is located in Ordos, Inner Mongolia, China. The terrain of the mining area is gentle, with a dip angle of less than 10°. At present, the 3-2 coal in the main coal seam has been mined out, and only the 4-2 coal under the industrial square, Baofu highway, and boundary coal pillar remain; the 5-2 coal is a thin seam with an average recoverable thickness of 1.24 m. Thus, a large quantity of coal gangue will be produced once mining commences. To extend the service life of the coal mine, improve the resource recovery rate, and digest the abandoned coal gangue, the short-wall PCMCB was considered for adoption for the pressed coal under the industrial square. The buried depth of the 4-2 coal is about 60 m, with a natural thickness of 0.70–3.45 m (average value: 2.57 m). The specific information and mechanical parameters of the coal seam roof and floor are shown in Table 5.

5.2 Determination of key parameters

First, the roof load q was determined, and the theoretical value of the ultimate span of rock beam damage was obtained according

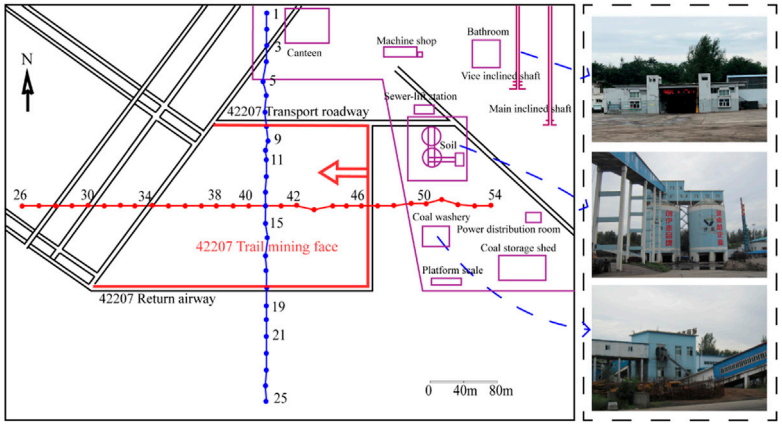


FIGURE 8
Layout of the surface subsidence measurement points and distribution of surface features.

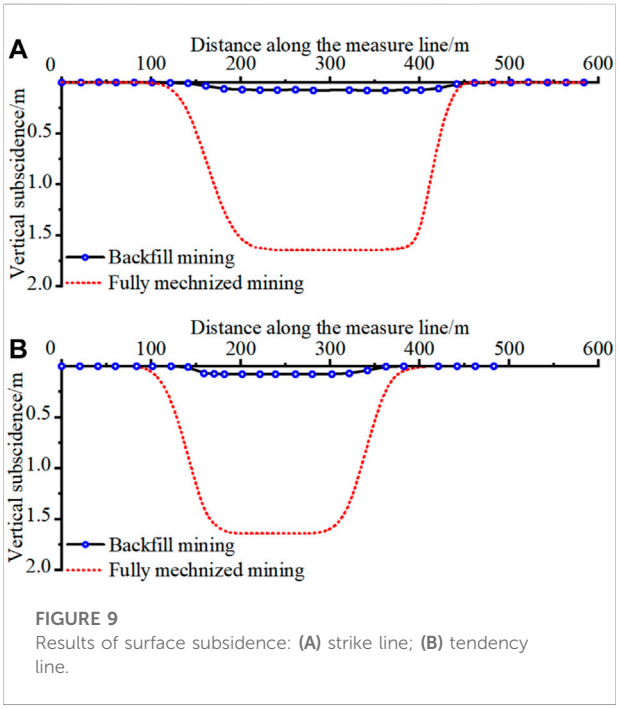


FIGURE 9
Results of surface subsidence: **(A)** strike line; **(B)** tendency line.

to the mechanical model of a simply supported beam. The width of the branch roadway was then selected as 5 m from engineering experience. The length of the branch roadway was selected as 200 m based on the lowest cost per ton of coal and field conditions. The design of the late strength of the backfill body was obtained according to Eqs 11–14. Combined with the mechanical model of the backfill body and Bieniawski formula, the late strength of the backfill body should be more than 2.62 MPa. Based on the design strength of the backfill body and backfill material proportions, the reasonable composition of cement: fly ash: gangue: water was determined as 1:1.82:4.17:1.78. To ensure that the deformation is within the first level of building damage, the critical filling rate of the goaf should be greater than 90% based on the equivalent mining height theory. The above data are summarized in Table 6.

5.3 Backfill process

The pattern of building the backfill plant on the ground was adopted in the Dingjiaqu coal mine, and the physical effect diagram is as shown in Figure 6. There are two stages in the backfill process. The first stage involves paste preparation, where the raw materials are weighed by proportion and then transported to the mixer through belt conveyors and pipelines for complete mixing. The second stage involves pipeline pumping, where the prepared paste is carried to the storage hopper and then transported to the underground working face through backfill pipelines.

The 42207 working face was mined as shown in Figure 1. After each of the branch roadways were mined, the upper and lower exits of the branch roadways were sealed with paste baffles first, and single hydraulic props were installed on the outside to strengthen the supports (Figure 7); finally, the paste was injected from the upper exit.

5.4 Measurement of surface subsidence

In line with the layout of the 42207 mining area and distribution of the surface features, third-class leveling was adopted for deformation monitoring as it has sufficient for subsidence monitoring. A strike observation line and a tendency observation line were arranged above the 42207 mining area, as shown in Figure 8. There were

TABLE 7 Technical and economic comparisons of several backfill technologies.

Backfill technology	Filling rate (%)	Cost per ton ($US)	Annual yield (million ton)	Applicable conditions	Interference
Solid backfill	60–85	6–10	0.3–0.5	Low sinking requirements	Large
Full mechanized mining and paste backfill	>95	12–16	0.2–0.4	Nearly horizontal coal seam	Large
PCMCB	>95	10–13	0.3	Small limitation	Small
High water backfill	80–85	18–22	0.2	Lack of solid waste	Large

25 observation points numbered 1 to 25 and set along the north–south observation line at distances of 20 m for a total length of 500 m; then, 29 observation points numbered 26 to 54 were set along the east–west observation line at distances of 20 m for a total length of 560 m.

The monitoring results (Figure 9) show that the subsidence could be reduced dramatically by backfilling and that the integral surface subsidence was small, with a maximum subsidence of 77 mm. The overall subsidence of the industrial square was uniform. Large inclination and horizontal deformations were observed at the edges of the subsidence basin as the coal seam is buried shallowly; accordingly, the inclination and horizontal deformation values changed rapidly. The maximum value of inclination was 1.99 mm/m; the maximum value of horizontal deformation was 0.56 mm/m; the maximum value of curvature deformation was 0.07×10^{-3} /m. Taking into account the errors between the forecast software and actual measurements, the observations basically show the law of surface deformation; a comprehensive analysis shows that the surface movement and deformations after backfilling meet the requirements of the building fortification level.

6 Discussion

To further reveal the advantages and disadvantages of the proposed short-wall PCMCB technology, a comparison between similar technologies is given in Table 7. Considering the characteristics of the western mining area, the solid backfill is less effective for controlling the surface deformations, and the high water backfill cannot handle a large amount of coal gangue (Hu et al., 2020). The surface control effect of the full mechanized and paste backfill is good, but its mining and backfilling interferences are serious, geological conditions of the working face are poorly adapted (Zhu, 2021), and the total investment in the full mechanized and paste backfill is generally more than twice that of the short-wall PCMCB, with the production capacity and tons of coal filling cost being similar (Zhu et al., 2017). Short-wall PCMCB also realizes full separation of the mining and backfilling, and the working face arrangement of this backfill method is flexible without being limited by the thickness of the coal seam; it also has strong applicability to the exploitation of overlaid and boundary coal resources (Lu et al., 2017; Li et al., 2022).

7 Conclusion

In the practice of three under mining, it is necessary to pay attention to improving the resource recovery rate and controlling the surface subsidence. Backfill mining not only achieves excellent recovery of the remaining coal resources but also reduces environmental damage effectively. To solve the problem of coal compressed under the industrial square in western China, the short-wall PCMCB scheme is proposed. The key to short-wall PCMCB is in the reasonable design of the width, length of the branch roadway, and strength of the backfill body. Based on the stress state of the branch roadway roof, it can be simplified as a simply supported beam, and the width of the branch roadway can be designed according to the bending and shear failure modes; the length of the branch roadway can be designed in accordance with the lowest cost per ton of coal; the late strength is calculated according to the strength design model of the backfill body. The filling rate is the standard for assessing whether the surface movement and deformation meet the requirements of building protection, which is obtained by the probability integral method. The selection and proportion of backfill materials are other important factors in the short-wall PCMCB; a reasonable proportion of the materials affords sufficient strength to the backfill body while maintaining backfill cost and digesting as much gangue as possible. The field application of the short-wall PCMCB in the Dingjiaqu coal mine shows that the subsidence could be reduced dramatically by backfilling. The integral subsidence of the industrial square was uniform, and the maximum subsidence values (77 mm) were far lower than the deformation extremes based on the requirements of the first level of building damage. This application proves that the short-wall PCMCB has good effects on the exploitation of the overlaid resources in western China.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, ZM and YD; methodology, YD; software, ZM and XZ; validation, XZ, FJ, YF, and CT; formal analysis, XZ; investigation, ZM; resources, XZ; data curation, ZM; writing—original draft preparation, ZM and YD; writing—review and editing, XZ; visualization, YD; supervision, XZ; project administration, XZ; funding acquisition, XZ. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (no. 52104203), Shandong Province Natural Science Foundation Project (no. ZR2020QE128), National Key R&D Program of China (no. 2018YFC0604705), SDUST Research Fund (no. 2018TDJH102), and National Natural Science Foundation of China (nos. 51574159, 52074251, 51974172, 51974173, 92058211, and 51804179).

Acknowledgments

The authors would like to thank the Dingjiaqu coal mine for providing field measurement data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, editors, and reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

References

Bai, E., Guo, W., Tan, Y., and Yang, D. (2018). The analysis and application of granular backfill material to reduce surface subsidence in China's northwest coal mining area. *PLoS One* 13 (7), e0201112. doi:10.1371/journal.pone.0201112

Belem, T., and Benzaazoua, M. (2007). Design and application of underground mine paste backfill technology. *Geotech. Geol. Eng.* 26 (2), 147–174. doi:10.1007/s10706-007-9154-3

Bieniawski, Z. T., and Van Heerden, W. (1975). The significance of *in situ* tests on large rock specimens. *Int. J. Rock Mech. Min. Sci. Geomechanics Abstr.* 12 (4), 101–113. doi:10.1016/0148-9062(75)90004-2

Chang, Q., Sun, Y., Leng, Q., Liu, Z., Zhou, H., and Sun, Y. (2021). Stability analysis of paste filling roof by cut and fill mining. *Sustainability* 13 (19), 10899. doi:10.3390/su131910899

Chi, X., and Liu, M. (2012). The stability analysis of stope roof based on beam model. *China Min. Mag.* 21 (08), 113–115.

Fan, L., and Ma, X. (2018). A review on investigation of water-preserved coal mining in Western China. *Int. J. Coal Sci. Technol.* 5 (4), 411–416. doi:10.1007/s40789-018-0223-4

Feng, G., Zhang, Y., Qi, T., and Kang, L. (2020). Status and research progress for residual coal mining in China. *J. China Coal Soc.* 45 (01), 151–159. doi:10.13225/j.cnki.jccs.YG19.1280

Gere, J. M., and Goodno, B. J. (2012). *Mechanics of materials*. Boston: Cengage learning.

Han, S., Chen, H., Long, R., and Cui, X. (2018). Peak coal in China: A literature review. *Resour. Conservation Recycl.* 129, 293–306. doi:10.1016/j.resconrec.2016.08.012

Hu, B., Liu, P., Cui, F., and Wang, H. (2020). Review and development status of backfill coal mining technology in China. *Coal Sci. Technol.* 48 (09), 39–47. doi:10.13199/j.cnki.cst.2020.09.004

Jia, K., and Feng, G. (2012). Application and prospect of super-high-water packing material in mining engineering. *Coal Sci. Technol.* 40 (11), 6–9+23. doi:10.13199/j.cnki.cst.2015.01.002

Jiang, N., Wang, C., Pan, H., Yin, D., and Ma, J. (2020). Modeling study on the influence of the strip filling mining sequence on mining-induced failure. *Energy Sci. Eng.* 8 (6), 2239–2255. doi:10.1002/ese3.660

Kang, H., Xu, G., Wang, B., Wu, Y., Jiang, P., Pan, J., et al. (2019). Forty years development and prospects of underground coal mining and strata control technologies in China. *J. Min. Strata Control Eng.* 1 (02), 7–39. doi:10.13532/j.jmsce.cn10-1638/td.2019.02.002

Koohestani, B., Belem, T., Koubaa, A., and Bussière, B. (2016). Experimental investigation into the compressive strength development of cemented paste backfill containing Nano-silica. *Cem. Concr. Compos.* 72, 180–189. doi:10.1016/j.cemconcomp.2016.06.016

Li, S., Zhao, Z., Yu, H., and Wang, X. (2021). The recent progress China has made in the backfill mining method, Part II: The composition and typical examples of backfill systems. *Minerals* 11 (12), 1362. doi:10.3390/min11121362

Li, Y., Lu, B., Yang, S., Lin, H., Xu, B., and Liu, C. (2022). Cemented backfilling mining technology with continuous mining and continuous backfilling method for underground coal mine and typical engineering cases. *J. China Coal Soc.* 47 (03), 1055–1071. doi:10.13225/j.cnki.jccs.2021.1612

Lin, H., Yang, R., Lu, B., Li, Y., Fang, S., Fan, Z., et al. (2021). Overlying strata movement law of continuous mining and continuous backfilling cemented-fill mining. *Environ. Earth Sci.* 80 (20). doi:10.1007/s12665-021-09993-w

Liu, K., Xu, J., and Miao, X. (2007). *Short-wall mining technology and its application*. Beijing: Coal Industry Press.

Liu, J., Li, X., and He, T. (2020). Application status and prospect of backfill mining in Chinese coal mines. *J. China Coal Soc.* 45 (01), 141–150. doi:10.13225/j.cnki.jccs.YG19.1063

Liu, Y. F., Wu, X. H., Zhu, T., Wang, X. J., Zhang, G. Y., and Wang, Z. G. (2022). Influence of mechanical properties of filling paste on overlying strata movement and surface settlement. *Shock Vib.* 2022, 1–9. doi:10.1155/2022/4687200

Lu, B., Zhang, X., Li, F., Zhang, B., and Pang, Z. (2017). Study and application of short-wall gangue cemented backfilling technology. *J. China Coal Soc.* 42 (S1), 7–15. doi:10.13225/j.cnki.jccs.2016.1416

Qian, M., Miao, X., and Xu, J. (1996). The key strata theory study in rock strata control. *J. China Coal Soc.* (03), 2–7.

Qu, Q., Yao, Q., Li, X., and Rong, T. (2010). Key factors affecting control surface subsidence in backilling mining. *J. Min. Saf. Eng.* 27 (04), 458–462.

Shao, X., Liu, E., Cai, X., Wu, J., and Zhang, J. (2018). Experimental study on stability of pillar in upward mining in pillar mined out area. *Coal Eng.* 50 (04), 5–9.

Sun, X., Qiang, H., Shi, X., and Chang, Q. (2017). Mixture ratio optimization of mine gangue paste filling materials. *Saf. Coal Mines* 48 (10), 69–72+75. doi:10.13347/j.cnki.mkaq.2017.10.018

Wang, L., Wei, T., Li, N., Chi, S. S., Jiang, C., and Fang, S. Y. (2018). Research on probability integration parameter inversion of mining-induced surface subsidence based on quantum annealing. *Environ. Earth Sci.* 77 (21). doi:10.1007/s12665-018-7927-z

Xiao, W., Chen, W., and Deng, X. (2021). Coupling and coordination of coal mining intensity and social-ecological resilience in China. *Ecol. Indic.* 131, 108167. doi:10.1016/j.ecolind.2021.108167

Xie, H., Duan, F., Zhou, H., and Zhao, G. (1998). Recent development of theory and analysis methods of strip pillar stability. *China Min. Mag.* (05), 37–41.

Yan, H., Zhang, J., Zhang, S., and Zhou, N. (2018). Physical modeling of the controlled shaft deformation law during the solid backfill mining of ultra-close coal seams. *Bull. Eng. Geol. Environ.* 78 (5), 3741–3754. doi:10.1007/s10064-018-1335-1

Yang, Z., Li, W., Pei, Y., Qiao, W., and Wu, Y. (2018). Classification of the type of eco-geological environment of a coal mine district: A case study of an ecologically fragile region in western China. *J. Clean. Prod.* 174, 1513–1526. doi:10.1016/j.jclepro.2017.11.049

Yang, Z., Li, W., Li, X., and He, J. (2019). Quantitative analysis of the relationship between vegetation and groundwater buried depth: A case study of a coal mine district in western China. *Ecol. Indic.* 102, 770–782. doi:10.1016/j.ecolind.2019.03.027

Zhang, Q., and Wang, X. (2007). Performance of cemented coal gangue backfill. *J. Cent. South Univ. Technol.* doi:10.1007/s11771–007–0043–y

Zhang, J., Li, M., Taheri, A., Zhang, W., Wu, Z., and Song, W. (2019a). Properties and application of backfill materials in coal mines in China. *Minerals* 9 (1), 53. doi:10.3390/min9010053

Zhang, Q., Zhang, J. X., Wu, Z. Y., and Chen, Y. (2019b). Overview of solid backfilling technology based on coal-waste underground separation in China. *Sustainability* 11 (7), 2118. doi:10.3390/su11072118

Zhang, Y., Liu, Y., Lai, X., Gao, J., and Wu, Y. (2021a). Physical modeling of the controlled water-flowing fracture development during short-wall block backfill mining. *Lithosphere* 2021 (4). doi:10.2113/2021/2860087

Zhang, Y., Wang, B., Zhang, X., and Li, F. (2021b). Forty years' development and future prospect on mechanized short-wall mining technology with continuous miner in China. *J. China Coal Soc.* 46 (01), 86–99. doi:10.13225/j.cnki.jccs.2020.1632

Zheng, B., Zhou, H., and He, R. (2006). Experimental research on coal gangue paste filling material. *J. Min. Saf. Eng.* (04), 460–463.

Zhu, L. (2021). Application and practice of solid waste multi-state filling technology in ecological fragile area of Western China. *Min. Saf. Environ. Prot.* 48 (04), 81–86. doi:10.19835/j.issn.1008-4495.2021.04.015

Zhu, W., Xu, J., Lai, W., and Wang, Z. (2007). Research of isolated section-grouting technology for overburden bed separation space to reduce subsidence. *J. China Coal Soc.* (05), 458–462.

Zhu, X., Guo, G., Zha, J., Chen, T., Fang, Q., and Yang, X. (2016). Surface dynamic subsidence prediction model of solid backfill mining. *Environ. Earth Sci.* 75 (12). doi:10.1007/s12665-016-5817-9

Zhu, W., Xu, J., Xu, J., Chen, D., and Shi, J. (2017). Pier-column backfill mining technology for controlling surface subsidence. *Int. J. Rock Mech. Min. Sci.* 96, 58–65. doi:10.1016/j.ijrmms.2017.04.014

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 17 January 2023
DOI 10.3389/feart.2022.1009855

OPEN ACCESS

EDITED BY
Wei Zhao,
Institute of Mountain Hazards and Environment (CAS), China

REVIEWED BY
Qiang Zou,
Institute of Mountain Hazards and Environment (CAS), China
Wang Lei,
Anhui University of Science and Technology, China

*CORRESPONDENCE
Long Yang,
yanglong1938@cug.edu.cn

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 02 August 2022
ACCEPTED 03 November 2022
PUBLISHED 17 January 2023

CITATION
Dai Z, Yang L, Zhang N, Zhang C, Zhang Z and Wang H (2023), Deformation characteristics and reactivation mechanism of an old landslide induced by combined action of excavation and heavy rainfall. *Front. Earth Sci.* 10:1009855. doi: 10.3389/feart.2022.1009855

Deformation characteristics and reactivation mechanism of an old landslide induced by combined action of excavation and heavy rainfall

Zhenwei Dai[1], Long Yang[1,2]*, Nan Zhang[3], Chenyang Zhang[2], Zhihua Zhang[4] and Heng Wang[4]

[1]Wuhan Center, China Geological Survey (Central South China Innovation Center for Geosciences), Wuhan, China, [2]Faculty of Engineering, China University of Geosciences, Wuhan, China, [3]China Institute of Geo-Environment Monitoring, Beijing, China, [4]Hydrogeology and Engineering Geology Team of Chongqing Bureau of Geology and Minerals Exploration, Chongqing, China

The reactivation mechanism of old landslides has been extensively studied from building load, erosion of the slope toe, heavy rainfall, and slope cutting for existing research. However, previous research on the reactivation of old landslides pays little attention to the combined action of engineering disturbance and heavy rainfall is rarely studied. This paper describes an old landslide in Wushan County, Chongqing, China, that was reactivated in August 2019 due to engineering disturbance and heavy rainfall. The deformation of the old landslide was first observed in 2007 and 2008, resulting from excavation and rainfall, respectively, and remained stable for about 11 years after treatment. In August 2019, the landslide was reactivated by slope cutting and damaged anti-sliding piles at the toe, and entered a state of imminent sliding due to the concentrated heavy rainfall events that occurred from October 4 to 22, 2019. In order to reveal the deformation features and reactivation mechanism of the landslide, field investigations, drilling activities and monitoring were performed. The results showed that tectonic effects and the stratigraphic lithology were the main reasons for the formation of the old Dashuitian landslide. The cut slope and damaged anti-sliding piles at the toe of the landslide provided the sliding space and reduced the anti-sliding force, and therefore resulted in the reactivation of the landslide. Continuous intense rainfall increased the weight of the landslide, decreased the mechanical properties and increased the pore water pressure of the weak interlayer, which accelerated the deformation rate. Therefore, 1.5 million m^3 of rock and soil masses slid along the weak interlayer under the action of gravity, threatening the safety of Wuliu Road, Ring Road, National Road G42 and the Wu-Da Expressway. Our research provides a theoretical basis for reducing the hazard of similar engineering projects involving slopes.

KEYWORDS

old landslide, engineering disturbances, rainfall, reactivation mechanism, deformation characteristics

1 Introduction

In recent years, hundreds of old landslides in China have been reactivated under a combination of internal and external geological dynamics and have resulted in major property damage and severe casualties (Yin et al., 2010; Guo et al., 2019). Examples include the Huangtupo landslide in Badong Town (Fourniadis et al., 2007), the Sanmendong landslide in Zigui County (Sun et al., 2016; Huo et al., 2019), the Liangshuijing landslide in Yunyang County (Xiao et al., 2013; Yin et al., 2016; Li et al., 2020) and the Baiyangwan landslide in Wushan Town (Zhang et al., 2020).

The reactivation of an old landslide is usually the result of multiple factors, among which heavy rainfall and disturbance from engineering projects are considered to be the most important (Cruden and Varnes 1996; Chen et al., 2014). Many old landslides have reactivated under the combined action of human engineering activities and concentrated rainfall. Zhang et al. (2020) reported that the old Baiyangwan landslide showed long-term creep due to the combined effects of building load, open excavation and heavy rainfall. By analyzing the deformation features of the old Outang landslide, Dai et al. (2016) reported that erosion of the slope toe by the Yangtze River, concentrated rainfall and reservoir impoundment were the main inducing factors leading to the reactivation of the landslide. Huang et al. (2020) investigated the reactivation mechanism of the old Dalixi landslide and found that excavation and rainfall together promoted the continuous deformation of the landslide. Due to seasonal heavy rainfall, overloading and slope cutting, the old Jiaju landslide in Danba county was reactivated and has undergone progressive retrograde failure since July 1992 (Li et al., 2008; Yin et al., 2010). With respect to old landslides, due to the existence of the old sliding zone and old cracks, the disturbance caused by excavation and heavy rainfall is more obvious and more likely to lead to landslide instability (Zhang et al., 2020). But how the combined action of the two factors affects the stability of an old landslide is little known.

The Dashuitian landslide, occurred in Wushan County, Chongqing, China, provided an opportunity to study reactivation mechanism of an old landslide induced by excavation and rainfall. Due to excavation that occurred in August 2019, the landslide reactivated and began to slowly slide, and professional monitoring was initiated on 30 August 2019. In October 2019, continuous and concentrated rainfall events promoted the deformation rate, and the landslide entered a state of imminent sliding. The total volume of the activated zone was about 150 $m^3 \times 10^4$ m^3, the average thickness was about 25 m, and the main sliding direction was about 7°. From August to December 2019, the total accumulated displacement of the Dashuitian landslide was about 150 mm. The landslide directly threatened the safety of Wuliu Road, Ring Road, G42 and the Wu-Da Expressway. The potential economic losses exceed 100 million CNY.

Based on detailed field investigations, drilling activities and field monitoring data, this study presents the particular characteristics of the Dashuitian landslide and its deformation features. The triggering factors were explored, and the reactivation mechanism was preliminarily analyzed. The effects of slope cutting and continuous intense rainfall on the reactivation of the old landslide were also examined. Considering that a large number of old landslides in China have been triggered by engineering disturbances and concentrated rainfall, we believe this study will provide useful guidelines and suggestions for the prevention and control of similar landslide events.

2 The Dashuitian landslide

2.1 General description

The Dashuitian landslide is located in Wushan County, Chongqing, China, which has undergone rapid development. The severe deformation of the landslide occurred in August 2019 (Figures 1A–C). The topography of the landslide area is high in the south and low in the north. The rear part of the Dashuitian landslide has an elevation of 540 m compared to 460 m at the leading edge, which is a height difference of 80 m. The angle of the slope is approximately 20°–25°. The G42 Expressway is located at approximately 80 m outside the northern boundary of the landslide. The leading edge of the landslide is a multistage stepped slope with an average angle of about 35°, which was excavated during the Wu-Da Expressway Interchange construction (Figure 1C). Ring Road is located at the middle-rear part of the landslide (Figure 1B). The rear edge of the landslide is bounded by Wuliu Road (Figure 1A). In addition, a gully has developed along the western edge of the landslide (Figures 1A, 2).

The landslide area is located at the south east wing of the Baishuihe syncline, which is the secondary fold of the Wushan syncline. The strata in the study area are monoclinal in structure with a northward dip, and the geological units, from top to bottom, include Quaternary landslide deposits (Q_4^{del}), and argillaceous limestone and mudstone of the Middle Triassic Badong Formation (T_2b^2) (Figure 3). The study area has a monsoon-influenced warm and humid subtropical climate with an average annual precipitation of 1049.5 mm. The rainy season mainly occurs from May to September, while the dry season occurs from December to February.

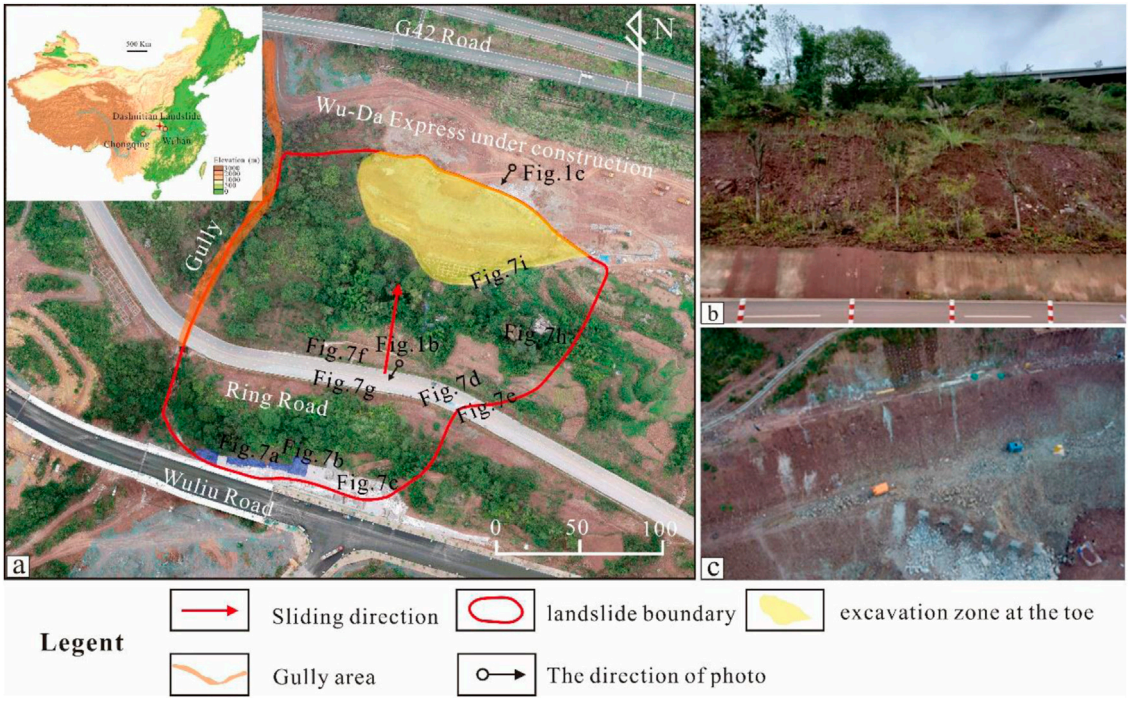


FIGURE 1
Photos of **(A)**—location and overview of the Dashuitian landslide, **(B)**—sliding body at the middle-rear part of the landslide, and **(C)**—excavation zone and broken anti-sliding piles at the toe of the landslide damaged during excavation.

2.2 Hydrogeological conditions

The study area is a sloping terrain with good drainage conditions, and there is no surface water in the area. Based on the occurrence and migration conditions, the groundwater can be divided into pore water of loose rock and fissure water of bedrock, and mainly flows from the southeast to the northwest. According to the investigation and drilling data, groundwater in the flat areas of the south and north boundaries of the Dashuitian landslide is abundant, while in the landslide area the groundwater is deep with no evident springs (Figures 2, 3). The pumping test results of the Zk6 and Zk10 boreholes showed that the permeability coefficients of the rock and soil masses range between 0.145 and 0.167 m/d, corresponding to a medium permeability layer.

3 Material composition

3.1 Sliding mass

The sliding mass consists of Quaternary landslide deposits (Q_4^{del}) and argillaceous limestone of the Middle Triassic Badong Formation (T_2b^2). The composition of the sliding mass varies greatly in the vertical direction. The upper part is a thin layer of gravel soil, and the lower part is a thick layer of broken limestone (Figure 3). Based on remote sensing images (Figure 1), the horizontal length of failed material is about 250 m, the maximum width is about 240 m, and the plane distribution area is about 6×10^4 m^2. In addition, the field investigation and drilling data revealed that the sliding direction of the landslide is 7°, the average thickness of the sliding body is about 25 m, and the unstable volume is about 150 m^3 × 10^4 m^3.

3.2 Sliding zone

The sliding zone occurs at depths of 7.5 (Zk6) to 32.9 m (Zk1) and is the contact zone between the argillaceous limestone and mudstone of the Middle Triassic Badong Formation (T_2b^2) with a thickness of about 0.8–1.3 m (Figures 4A–C). The dip angle of the sliding surface is about 17–25° and exhibits characteristics of deep burial depth in the southwest and shallow burial depth in the northeast. There is a weak interlayer in the sliding zone, which is composed of broken argillaceous limestone and mudstone clasts (Figures 3, 4B,C). Many polished surfaces, scratch, and slickensides were observed in the sliding zone (Figures 4D–G). Based on the results of direct shear tests, the mechanical properties of the sliding zone materials are poor, especially in the saturated state (Table 1).

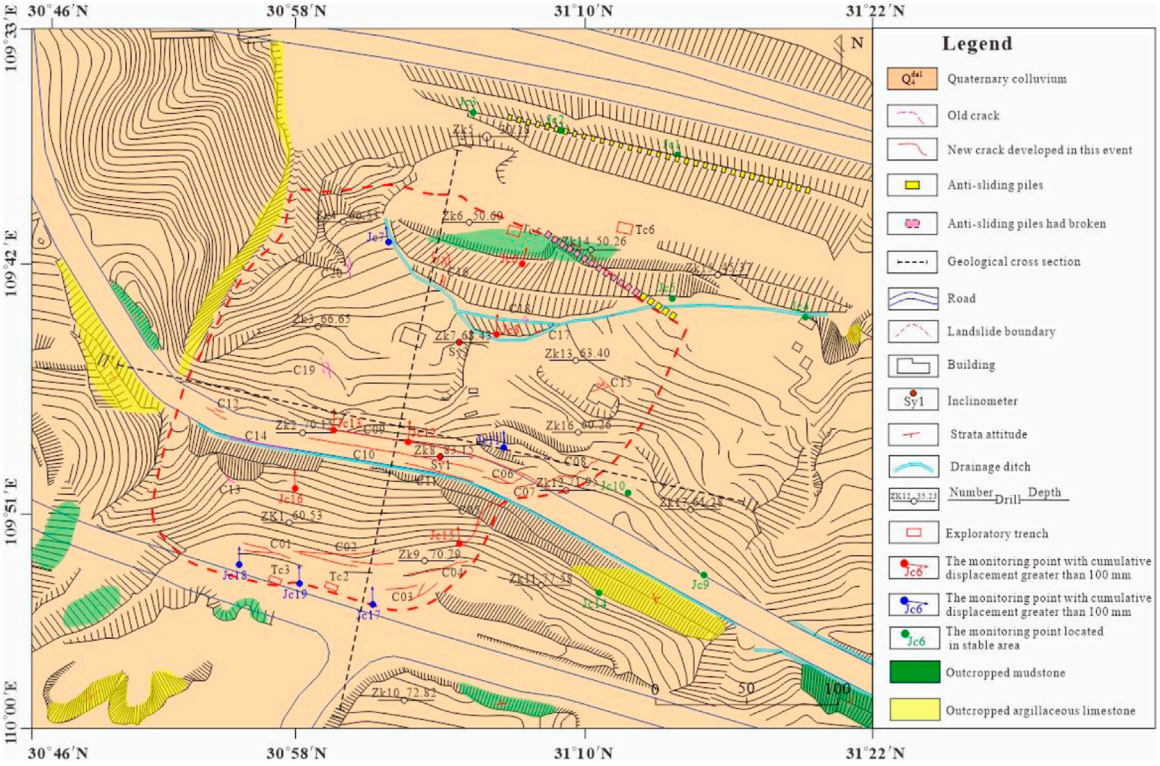


FIGURE 2
Layout map of the landslide area, showing the stratigraphic boundaries, landslide boundary and the location of boreholes, monitoring points, inclinometers, exploratory trenches and cracks. C13, C14, C18, C19, and C20 are old cracks induced by previous failure events, and the others are new cracks caused by this failure event.

3.3 Sliding bed

The sliding bed is composed of the mudstone, and the attitude of the strata is 330° with a dip of 25°. The mudstone is purple, features a medium-thick layered structure and appears relatively intact compared with the overlying argillaceous limestone. However, the mudstone near the sliding zone is highly weathered due to softening by groundwater (Figure 5).

4 Human engineering activities

In the past 20 years, with rapid economic development, many human engineering disturbances have taken place in the landslide area, including along Wuliu Road at the rear part of the landslide, Ring Road at the middle–rear part, the Wu-Da Expressway at the toe and the G42 Expressway at the leading edge. Both Wuliu Road and Ring Road cross the sliding body, forming subgrade slopes with lengths of 450 and 480 m, and heights of 1–12 and 15–30 m, respectively (Figure 1). The G42 Expressway is located at approximately 80 m outside the northern boundary of the landslide and forms a subgrade slope with a length of 450 m, a height between 25 and 35 m and an average slope angle of 50°. 18 anti-sliding piles, with a length of 28–30 m, a cross section of 2 m×3 m and a buried depth of 15–25 m, were established at the landslide toe to prevent its deformation caused by heavy rainfall in August 2008. In August 2019, 30 $m^3 \times 10^4$ m^3 of rock and soil masses was excavated, and 13 anti-sliding piles were destroyed at the toe of the landslide during the construction of the Wu-Da Expressway. As a result of the excavations, three high-steep slopes were formed at the toe with heights between 25 and 35 m and a slope angle of 35° (Figure 1C).

5 Landslide deformation characteristics

5.1 Field investigation

Five typical old cracks (C13, C14, C18, C19, C20) were developed prior to this event and were revealed during a detailed field investigation (Figures 2, 6). C13 is a tension crack of about 30 m in length, 5–20 cm in width and 50 cm in depth, and is located

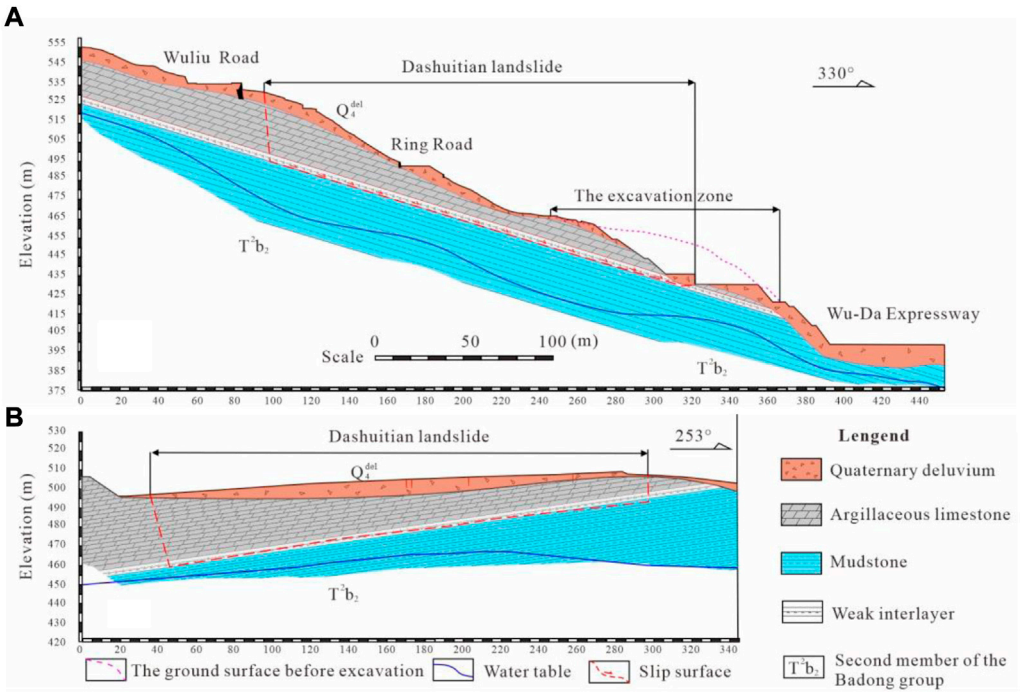


FIGURE 3
Cross sections (sections shown in Figure 2) of the Dashuitian landslide. **(A)**—Geological I-I′ cross-section; **(B)**—Geological II-II′ cross-section. The blue lines are the groundwater levels.

at the rear part of the Dashuitian landslide (Figure 6). C14 is also a tension crack that developed at the middle-rear part of the landslide with a length of 2 m and width of 2–4 cm (Figure 6). C18 is a shear crack located at the leading part of the landslide (Figure 6). The investigation indicated that these cracks are historical cracks that were induced by the excavation of the G42 Expressway in 2007 and 2008.

The new deformation signs of the Dashuitian landslide that were identified during the field investigation mainly include cracks on the roads and ground surface (Figure 7). Deformation signs became evident after the heavy rainfall event of October 4 to 22, 2019, when about 228 mm of precipitation fell (Figure 8). The new deformation signs mainly developed at the leading part, rear edge, the middle-rear part and eastern edge of the landslide. C15, C16 and C17 developed at the leading part of the landslide, reaching a length and width of 7–10 m and 1–2 cm, respectively. The deformation signs at the rear edge developed along Wuliu Road and consist of several transverse tension cracks, such as C01 and C02. The length of these is 20–50 m, the width is 1–40 cm and the visible depth is 20–200 cm. C09–C12 are large tension cracks along Ring Road at the middle-rear part of the landslide with a length of 8–120 m. C03–C08 are feathery shear cracks, located along the eastern edge of the landslide, with a length of 8–40 m, width of 1–25 cm and depth of 1–50 cm. Overall, the greatest deformation mainly occurs at the middle-rear part of the Dashuitian landslide, while the deformation in the leading part is relatively minor (Figure 2).

5.2 GNSS-based monitoring data

After 30 August 2019, we installed GNSS (Global Navigation Satellite Systems) in the study area to monitor the ground deformation, as shown in Figure 2. Based on the results of the field investigation and the geomorphological features of the study area, the gully located at the western edge of the landslide was determined as the western boundary. Moreover, in order to further determine the northern, southern and eastern boundaries of the landslide, monitoring points JC1, JC2, JC3, and JC5 were installed on the northern side, JC17, JC18, and JC19 were developed at the southern edge, and JC4, JC9, JC10, and JC14 were situated along the eastern side. The monitoring results showed that the cumulative displacement of all 11 monitoring points was about 2–10 mm, and the deformation rate was only 0.02–0.09 mm/d from August 30 to 2 December 2019, indicating that the positions of the monitoring points are relatively stable (Figures 8A–C).

FIGURE 4
(A) is a weak interlayer in the sliding zone composed of broken mudstone and argillaceous limestone, which is revealed in Zk16, located at 17.5–18.6 m with a thickness of 1.1 m; **(B,C)** are argillaceous limestone and weathered mudstone, as revealed in Zk16, with respective depths of 17.7 m and 18.1 m; and **(D–G)** are revealed in Zk3, Zk7, Zk8 and Zk9, respectively, and contain many polished surfaces and slickensides.

TABLE 1 Shear strength parameters of Dashuitian landslide materials.

	Material	Natural state unit weight (kN/m^3)	Cohesive (MPa)	Internal friction angle (°)	Saturated state unit weight (kN/m^3)	Cohesive (MPa)	Internal friction angle (°)
sliding mass	Quaternary colluvium	20.3	0.021	16.43	20.9	0.017	13.9
	Argillaceous limestone	26.50	7.67	45.41	26.60	6.14	34.06
Sliding zone	Weak layer	20.4	0.022	16.47	20.9	0.018	14.24
Sliding bed	Mudstone	26.20	1.90	37.73	26.30	1.23	28.30

Regarding the activated zone, two typical stages in the deformation curves are shown in Figures 8D–F, which correspond perfectly with the concentrated rainfall events from October 4 to 22, 2019, with a total precipitation of 228 mm. Monitoring points JC6, JC7, and JC8, which are located at the open excavation zone in the leading part of the landslide, had a maximum deformation rate of 1.29 mm/day before the rainfall event and 1.61 mm/day after the rainfall

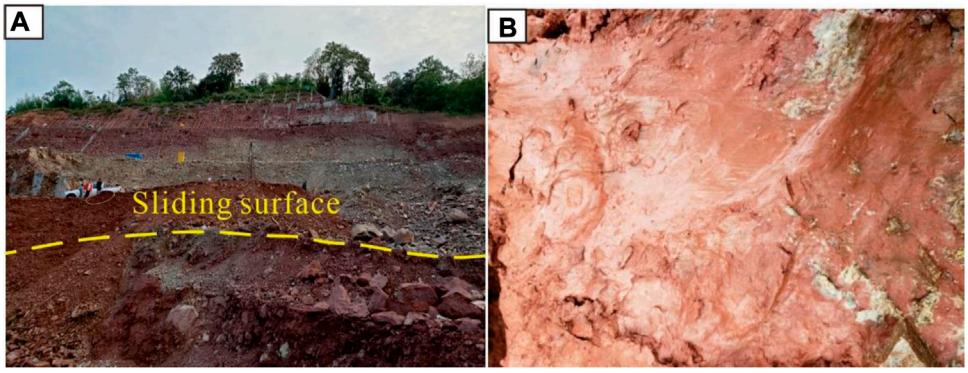


FIGURE 5
(A) is the sliding zone revealed by the excavation; **(B)** is the softened sliding zone exposed by TC5 (Figure 2) at a depth of 16 m.

FIGURE 6
The typical old cracks developed in 2007 and 2008 due to deformation of the old Dashuitian landslide.

event. The highest deformation rate of monitoring points JC11, JC12 and JC13, located at Ring Road in the middle part of the landslide, reached approximately 1.43 mm/day and 1.87 mm/day before and after the rainfall event, respectively. The highest deformation rate of monitoring points JC15 and JC16, located at the rear part of the landslide, was approximately 0.97 mm/day and 1.25 mm/day before and after the rainfall event, respectively. In addition, the total cumulative displacement in the leading (JC6, JC7, and JC8), middle (JC11, JC12, and JC13) and rear (JC15, JC16) parts of the landslide were about 85.9–134.7, 96.5–154.3, and 103.1–105 mm, respectively.

5.3 Subsurface deformation characteristics

Two inclinometers were used to measure the subsurface deformation, and their positions are shown in Figure 2. The range of the inclinometer used in this monitoring was ±15°, and the accuracy was 0.01 mm/500 mm.

The monitoring displacement curve of SY01 indicates that the displacement increased gradually from the bottom to the top (Figure 9). An obvious sliding surface was evident between 17 and 20 m. From October 9 to 2 December 2019, the cumulative horizontal displacement and deformation rate above the sliding surface reached 78 mm and 1.5 mm/day, respectively. Based on data from borehole Zk16 (Figure 4A), the weak interlayer was inferred at a depth of 16–20.5 m, which is the contact zone between the argillaceous limestone and mudstone. Given that scratches were apparent in the borehole core, the contact zone is considered to be a sliding surface.

The monitoring curve of SY02 shows an obvious sliding surface at a depth of 19–21 m, and the unstable rock and soil masses above the sliding surface moved as a whole, suggesting that the Dashuitian landslide has a well-defined deep sliding surface (Figures 3, 9). The cumulative horizontal displacement and deformation rate above the sliding surface reached 48 mm and 1.6 mm/day from October 20 to 2 November 2019. Based on information from borehole Zk7 (Figure 4E), the contact zone between the argillaceous

FIGURE 7
New deformation signs of the Dashuitian landslide. **(A,B)** are tensile cracks located at the rear part; **(C–E)** are shear cracks located at the eastern edge; **(F,G)** are large tensile cracks develop along Ring Road at the middle-rear part; **(H,I)** are tensile cracks located at the leading part.

limestone and mudstone is located at 16.7–19.2 m, and the sliding surface is interpreted to be located near this contact zone.

In conclusion, the locations of the sliding surface revealed by the two inclinometers (Figure 9) are nearly consistent with the contact zone between the argillaceous limestone and mudstone revealed by the boreholes (Figure 4).

6 Discussion

6.1 Influencing factors of the Dashuitian landslide

The reactivation mechanism of landslides has been extensively studied from building load (Zhang et al., 2020), erosion of the slope toe by river (Dai et al., 2016),

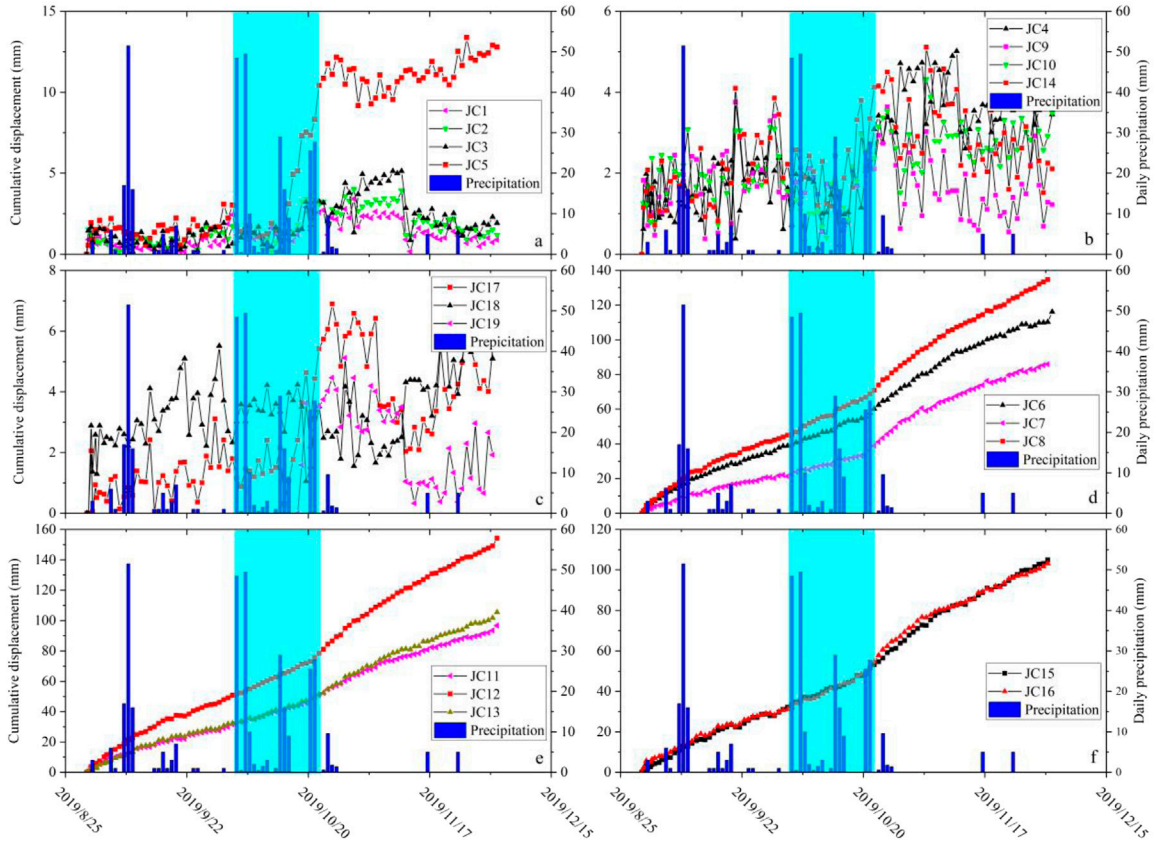


FIGURE 8
Monitoring results of the cumulative displacement and daily precipitation from August 30 to 2 December 2019. **(A–C)** are boundary control points; **(D)** is located at the open excavation zone in the leading part of the landslide; **(E)** is located along Ring Road; and **(F)** has been installed at the rear part of the landslide. The cyan section corresponds to the period from October 4 to 22, 2019, during which a continuous heavy rainfall event occurred, with a total precipitation of 228 mm.

heavy rainfall (Huang et al., 2020) and slope cutting (Li et al., 2008; Yin et al., 2010) for previous research. With respect to old landslides, the disturbance caused by excavation and heavy rainfall is more obvious and more likely to lead to landslide instability due to the existence of the old sliding zone and old cracks (Zhang et al., 2020). However, the combined action of engineering disturbance and heavy rainfall on the reactivation of old landslides is rarely studied. Under such scenarios, we mainly focus on the predisposing factors (including tectonic effect and formation lithology) and triggering factors (including rainfall and human engineering activities) of the Dashuitian landslide to study its reactivation mechanism. Our research may provide a theoretical basis for reducing the hazard of similar engineering projects involving slopes.

6.2 Predisposing factors

6.2.1 Tectonic effect

The study area is located in the northwest wing of the Wushan syncline where the Baishuihe syncline is developed. Due to the effects of regional tectonic action, the rock and soil masses in the landslide area are extremely broken (Figure 4) and easily infiltrated by rainwater and groundwater.

6.2.2 Formation lithology

Broken rock and soil masses compose the Quaternary colluvium layer of the Dashuitian landslide, which is rapidly infiltrated by rainfall and surface water due to its loose structure and good permeability. The underlying layer is argillaceous limestone, where joints and fractures are well developed due to tectonic compression (Chai et al., 2013; Yan et al., 2019). These

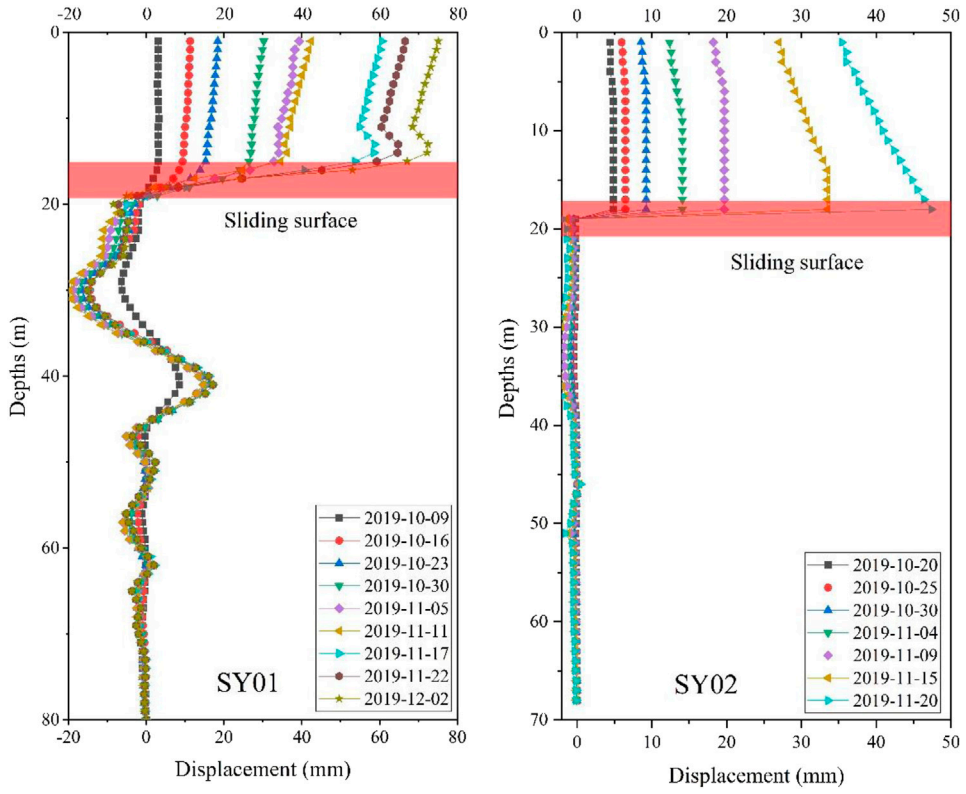


FIGURE 9
The cumulative displacement monitoring results of inclinometers SY01 and SY02.

joints and fractures provide channels for the rapid infiltration of rainwater. The bedrock is mudstone, which is relatively intact and has poor permeability compared with the upper argillaceous limestone. Therefore, rainwater can rapidly infiltrate to the contact surface between argillaceous limestone and mudstone, and form stagnant water in the sliding zone, especially during rainfall. A weak interlayer therefore presents near the contact surface, which is heavily weathered, has poor mechanical properties and is prone to slippage (Zhang et al., 2020).

6.3 Triggering factors

6.3.1 Human engineering activities

During the construction of the Wuliu and Ring roads, the slope body showed no obvious deformation signs. In 2007, due to the construction of the G42 Expressway, a large-scale excavation was carried out at about 80 m outside the toe of the landslide, causing apparent deformation for the first time. In order to control the deformation of the slope, 33 anti-sliding piles were established at the toe (Figure 2). However, under the influence of concentrated rainfall in July and August 2008, new deformation occurred. An additional 18 anti-sliding piles were therefore established at the slope toe (Figure 2). No obvious deformation had occurred since then. In August 2019, 30 $m^3 \times 10^4$ m^3 of rock and soil masses was excavated from the toe area during the construction of the Wu-Da Expressway, which provided the necessary conditions for reactivation. In addition, the anti-sliding force of the landslide was reduced due to damage to 13 of the anti-sliding piles at the toe (Figures 1C, 2), which triggered the overall slow deformation of the Dashuitian landslide during August 30 to 22 October 2019 (Figure 8).

6.3.2 Rainfall

Prior to the construction of the Wu-Da Expressway, the maximum daily precipitation recorded in Wushan County was 384.6 mm on 31 August 2014 (Yan et al., 2019), which is much higher than the total precipitation of 228 mm that fell from October 4 to 22, 2019 (Figure 8). The maximum recorded daily precipitation (346.8 mm) did not induce any significant deformation to the Dashuitian landslide. However, following the construction of the Wu-Da Expressway in August 2019, the lower rainfall amount promoted the deformation rate of the landslide, which is supported by the GNSS-based monitoring data collected after 22 October 2019 (Figure 8). Therefore, compared with engineering activities, the effect of rainfall was not a decisive factor that led to the reactivation of the Dashuitian landslide. The concentrated heavy

rainfall increased the pore water pressure and reduced the mechanical properties of the weak interlayer, increased the weight of the Dashuitian landslide, and finally promoted the deformation rate. Zhang, 2020.

6.4 Reactivation mechanism

The extensive excavation at the toe of the Dashuitian landslide provided the sliding space for reactivation. Meanwhile, the anti-sliding force of the landslide was significantly reduced due to damage to 13 of the anti-sliding piles located at the toe. Thus, in August 2019, the Dashuitian landslide slid slowly along the weak interlayer. Moreover, the sliding body is composed of loose Quaternary colluvium and argillaceous limestone broken by tectonic compression, which provide channels for the rapid infiltration of rainwater. The old cracks in particular, formed by the deformation of the old Dashuitian landslide, accelerated the infiltrate process. Rainwater can therefore quickly infiltrate the sliding body, increasing the water content of the rock and soil masses above the sliding zone, and increasing the sliding force of the landslide. Most importantly, the bedrock is composed of mudstone, which is relatively intact and has poor permeability compared with the upper argillaceous limestone. Therefore, the weak interlayer remained saturated during the heavy rainfall, which significantly increased its pore water pressure and reduced the shear strength. This dramatically reduced the anti-sliding force on the weak interlayer, thereby promoting the deformation rate of the Dashuitian landslide. Finally, the landslide entered a state of imminent sliding on 22 October 2019.

7 Conclusion

The Dashuitian landslide is an old landslide with a volume of 150 m^3 × 10^4 m^3, which reactivated in August 2019, and then entered a state of imminent sliding on 22 October 2019. The area of greatest deformation is mainly located at the middle-rear part of the landslide, while the deformation at the leading part is relatively minor. Based on the field investigation, monitoring data and drilling results, the reactivation mechanism of Dashuitian landslide could be deduced as follows. Tectonic effects and formation lithology are the main reasons for the formation of the old landslide. The slope cutting and the damaged anti-sliding piles at the landslide toe provided the sliding space, reduced the anti-sliding force and therefore resulted in the reactivation of the landslide. Moreover, the concentrated rainfall increased the weight of the Dashuitian landslide, reduced the mechanical properties and increased the pore water pressure of the weak interlayer and finally promoted the landslide to enter a state of imminent sliding. Our research may provide a theoretical basis for reducing the hazard of similar engineering projects involving slopes.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

The work was supported by funding from National key R&D plan (2018YFC1505401-3), Natural Science Foundation of Hubei Province (Grant No. 2020CFB352), a follow-up of the Geological Disaster Prevention and Control Project in the Three Gorges area (Grant Nos 000121 2019C C60 001 and 000121 2021C C60 001), and Chongqing Natural Science Key program: Key Technology for Treatment of Massive Hydraulic Landslide in Three Gorges Reservoir Area (Grant No. cstc2020jcyj-zdxmX0019).

Acknowledgments

The authors would like to thank Xiannian Jiang, Baibing Jiang, Dan Luo, and Peng Jin from 208 Geological Team, Chongqing, for their great assistance in the field investigation and providing monitoring data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Chai, B., Yin, K. L., Du, J., and Xiao, L. L. (2013). Correlation between incompetent beds and slope deformation at Badong town in the Three Gorges Reservoir, China. *Environ. Earth Sci.* 69, 209–223. doi:10.1007/s12665-012-1948-9

Chen, C. L., He, K., and Li, T. L. (2014). Research on the mechanism of the ancient landslide resurrection triggered by slope toe excavation. *Northwest. Geol.* 47 (1), 255–260. in Chinese.

Cruden, D. M., and Varnes, D. J. (1996). Landslide types and processes, special report, transportation research board. *Natl. Acad. Sci.* 247, 36–75.

Dai, Z. W., Yin, Y. P., Wei, Y. J., Lu, T., Luo, J. H., and Wang, Y. (2016). Deformation and failure mechanism of outang landslide in three gorges reservoir area. *J. Eng. Geol.* 24 (1), 44–55. In Chinese. doi:10.13544/j.cnki.jeg.2016.01.006

Fourniadis, I. G., Liu, J. G., and Mason, P. J. (2007). Landslide hazard assessment in the Three Gorges area, China, using aster imagery: Wushan–Badong. *Geomorphology* 84 (1), 126–144. doi:10.1016/j.geomorph.2006.07.020

Guo, C. B., Zhang, Y. S., Li, X., Ren, S. S., Yang, Z. H., Wu, R. A., et al. (2019). Reactivation of giant jiangdingya ancient landslide in zhouqu county, gansu province, China. *Landslides* 17, 179–190. doi:10.1007/s10346-019-01266-9

Huang, X. H., Yi, W., Gong, C., Huang, H. F., and Yu, Q. (2020). Reactivation and deformation mechanism of ancient landslides by excavation. *Chin. J. Geotechnical Eng.* 42 (7), 1276–1285. in Chinese. doi:10.1155/2021/6614180

Huo, Z., Huang, B., Wang, L., Lu, S. Q., and Fan, Y. (2019). Potential surge risk of Sanmendong landslide in three gorges reservoir area. *J. Catastrophology* 8 (1), 454–461. doi:10.1080/21642583.2020.1788469

Li, M. H., Zheng, W. M., Shi, S. W., and Xie, Z. S. (2008). The revival mechanism and stability analysis to Jiaju landslide of Danba county in Sichuan province. *J. Mt. Sci.* 26 (5), 577–582. in Chinese. doi:10.21203/rs.3.rs-555768/v1

Li, Y., Utili, S., Milledge, D., Chen, L. X., and Yin, K. L. (2020). Chasing a complete understanding of the failure mechanisms and potential hazards of the slow moving liangshuijing landslide. *Eng. Geol.* 281, 105977. doi:10.1016/j.enggeo.2020.105977

Sun, G., Zheng, H., Huang, Y., and Li, C. (2016). Parameter inversion and deformation mechanism of Sanmendong landslide in the Three Gorges Reservoir region under the combined effect of reservoir water level fluctuation and rainfall. *Eng. Geol.* 205, 133–145. doi:10.1016/j.enggeo.2015.10.014

Xiao, S. R., Lu, S. S., Guan, H. F., and Song, G. L. (2013). Study of geomechanical model of liangshuijing landslide in three gorges reservoir area. *Rock Soil Mech.* 34, 3534–3542. In Chinese.

Yan, G. Q., Yin, Y. P., Huang, B. L., Zhang, Z. H., and Zhu, S. N. (2019). Formation mechanism and characteristics of the jinjiling landslide in wushan in the three gorges reservoir region, China. *Landslides* 16 (11), 2087–2101. doi:10.1007/s10346-019-01234-3

Yin, Y. P., Huang, B. L., Wang, W. P., Wei, Y. J., Ma, X. H., Ma, F., et al. (2016). Reservoir-induced landslides and risk control in three gorges project on Yangtze River, China. *J. Rock Mech. Geotechnical Eng.* 8, 577–595. doi:10.1016/j.jrmge.2016.08.001

Yin, Y. P., Zheng, W. M., Liu, Y. P., Zhang, J. L., and Li, X. C. (2010). Integration of GPS with InSAR to monitoring of the Jiaju landslide in sichuan, China. *Landslides* 7 (3), 359–365. doi:10.1007/s10346-010-0225-9

Zhang, C. Y., Yin, Y. P., Dai, Z. W., Huang, B. L., Zhang, Z. H., Jiang, X. N., et al. (2020). Reactivation mechanism of a large-scale ancient landslide. *Landslides* 18 (1), 397–407. doi:10.1007/s10346-020-01538-9

Zhang, Z. (2020). Mechanism of the 2019 Yahuokou landslide reactivation in Gansu, China and its causes. *Landslides* 17 (1), 1429–1440. doi:10.1007/s10346-020-01384-9

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/feart.2023.1082854

OPEN ACCESS

EDITED BY
Lixin Wu,
Central South University, China

REVIEWED BY
Wei Shan,
Northeast Forestry University, China
Pengpeng Ni,
Sun Yat-sen University, China

*CORRESPONDENCE
Yu Wan,
wy0209@hhu.edu.cn

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 28 October 2022
ACCEPTED 13 February 2023
PUBLISHED 24 February 2023

CITATION
Gu J, Chen L, Wan Y, Teng Y, Yan S, Hu L, Wang J and Luo H (2023), Study on the electric resistance method in crack depth measurements.
Front. Earth Sci. 11:1082854.
doi: 10.3389/feart.2023.1082854

Study on the electric resistance method in crack depth measurements

Jiahui Gu[1,2], Liang Chen[1,2], Yu Wan[1,2]*, Yaozong Teng[1,2], Shufa Yan[1,2], Liang Hu[1,2], Jiahao Wang[1,2] and Hanxuan Luo[1,2]

[1]School of Civil Engineering and Transportation, Hohai University, Nanjing, China, [2]Key Laboratory of Geomechanics and Embankment Engineering, Hohai University, Nanjing, China

Obtaining geometric parameters, especially depth, and describing the morphological characteristics of cracks are of great significance to control engineering disasters and accidents caused by cracks. The electric resistance method is based on the principle of differences in electrical properties between cracks and soil, which could be used to measure the single crack depth at project sites. There exists an R_{min} value corresponding to a specific electrode distance d value at each R_f-d value obtained by laboratory experiments. Furthermore, a two-dimensional finite element model of soil with a single crack is established to carry out numerical simulation analysis considering the crack width W, crack depth D and complex crack conditions. The results reveal dynamic variation rules of soil resistance after crack development, and for each R_f-d value, the electrode distance d value corresponding to the R_{min} value is approximately equal to the crack depth D. In the range of the electric field, the offset and rotation of the crack have little effect, while the measurement results have a strong dependence on relocation movement. The regulation gives guidance to the inversion analysis of crack depth D at project sites and has been applied in crack depth measurements of an expansive soil slope. The electric resistance method as a proposed integrated approach is of great significance and brings new perspectives into the study of crack depth measurements for field applications.

KEYWORDS

crack depth, electric resistance method, numerical Simulation, field application, finite element analysis

1 Introduction

Due to global climate changes, continuous extreme aridity has increased substantially in East and South Asia, Africa, southern Europe and eastern Australia (Dai, 2011), imposing detrimental impacts on the regional environment, infrastructure construction and people's livelihoods. Influenced by aridity and desiccation, a large amount of water evaporates from the soil, resulting in soil shrinkage (Lakshmikantha, et al., 2012), crack development and changes in the soil physical and mechanical properties (Albrecht and Benson, 2001; Dyer, 2005; Hallett and Newson, 2011; Costa et al., 2013; Kodikara and Costa, 2013; Kong et al., 2017). The existence of cracks destroys the soil integrity and contributes to attenuation of the soil strength (Lozada et al., 2015; Kong et al., 2018). At the same time, the large crack networks provide many seepage channels for the migration of water (Yuan and Yin, 2004; Krisnanto et al., 2014; Li et al., 2017), leading to the deterioration of soil physical and mechanical properties (Chaduvula et al., 2017; Trabelsi et al., 2018). Therefore, cracks are a

direct or indirect incentive for many kinds of geotechnical, geological and geo-environmental problems. For instance, the formation of cracks leads to a decline in expansive soil slope stability and accelerates rainwater infiltration, which is prone to cause landslides (Yin et al., 2012; Xie et al., 2020; Zhang et al., 2021). Crack development may bring about a decrease in the foundation bearing capacity or an increase in the foundation soil compressibility, causing the inclination or even collapse of the construction (Ramana, 1993; Lozada et al., 2015). Cracks in dams are likely to generate dam bursts (Dyer et al., 2009; Khandelwal et al., 2013). In waste disposal storage, the cracks in the soil layer of the isolation zone contribute to the leakage of pollutants (Hewitt and Philip, 1999; Rayhani et al., 2007; Baram et al., 2013).

In recent years, increasing attention has been given to the problem of cracks in geotechnical and geological engineering (Baer et al., 2009; Shin and Santamarina, 2011; Hirobe and Oguni, 2016; Tollenaar et al., 2017; Vo et al., 2017; Wang et al., 2017; Pouya et al., 2018; Tang et al., 2018; Vo et al., 2018; Wang et al., 2018; Li et al., 2019; Ruan et al., 2021). Obtaining the geometrical parameters of the primary crack after development, especially the depth, can be used to control engineering disasters in a more accurate, efficient and economical way. For example, to control expansive soil slope landslide disasters by replacing cement-based materials or other treatment methods, it is urgent to determine the depth of slope failure. Otherwise, it will cause huge economic waste, and the treatment effect cannot be guaranteed (Cai and Li, 2018).

Scholars worldwide have used a series of direct and indirect methods to obtain the geometric parameters of cracks. Lu et al. (2002) defined crack damage variables based on CT data and analyzed the variation rule of crack damage variables with cumulative dry shrinkage. X-ray CT, as a non-destructive, non-invasive, real-time and 3D visualization method of the inner microstructures, has provided a promising way to observe the geometry of cracks in test specimens in real time and explore the mechanical behaviors and damage mechanism of geomaterials (Li et al., 2015; Yang et al., 2015; Cala et al., 2016; Kim et al., 2016; Yu et al., 2016; Zhao and Zhou, 2020). Stewart et al. (2012) used a horizontal recorder connected to a crack to record the water level in the pipe corresponding to volumetric changes in the crack. The methods above are capable of describing the crack geometric parameters dynamically, quantitatively and non-destructively, but the machines are expensive and demand high-quality operation confined to the laboratory. Moreover, scholars worldwide have carried out theoretical algorithm methods and established crack depth calculation models based on the mechanism of crack development and various mechanics, such as linear elastic mechanics and fracture mechanics (Morris et al., 1992; Chertkov, 2000; Zhen et al., 2006; Taghavi et al., 2010; Zhu et al., 2020). However, the predicted value of the crack depth obtained by theoretical calculations is the initial development crack depth, not the final development depth in the stable state. Meanwhile, theoretical calculations contain various assumptions, and there exists a large deviation between the predicted value and the actual value.

Compared with the methods above, the electrical resistivity method as a non-destructive technique has strong operability. Based on differences in the electrical properties between cracks and the soil, a series of tests in the laboratory and *in situ* measurements have been conducted with the electrical resistivity method to perform tomography scans, which proved the applicability of using electrical properties to express crack development and the geometrical shape (Samouëlian et al., 2003; Samouëlian et al., 2004; Ackerson et al., 2014; Jones et al., 2014; Gunn et al., 2015; Tang et al., 2018; An et al., 2019). Various approaches to invert 1D, 2D and 3D resistivity data have been published (Sultan et al., 2006; Pidlisecky et al., 2007; Park et al., 2009; Jones et al., 2012; Damavandi et al., 2022). Inversion of 2D resistivity data has been used to interpret geoelectrical resistivity data and construct 2D images of subsurface structures. The electrical resistivity imaging technique can be used to recognize underground structures by observing contrasts in the subsurface resistivity distribution (Sungkono et al., 2014; Neyamadpour, 2018; Fma et al., 2022). This method provides a new viewpoint for crack quantitative studies, and the operation requirements are relatively easy to satisfy, but the study is limited to laboratory experiments. Although the electrical resistivity imaging technique has been used to describe the crack distribution underground *in situ*, there still exists a problem of deviation, which cannot be ignored, owing to simplification in the complex inversion process.

Focusing on the primary crack, which is rather long and plays a decisive role in engineering safety, the study reported herein proposes a new method to carry out crack depth measurements for better field applications, i.e., the electric resistance method based on the theory that the soil resistance after the test current passes through the soil could reflect the internal structure of the soil mass indirectly. Laboratory experiments are conducted, and a two-dimensional finite element model of soil with a single crack is established to carry out numerical simulation analysis considering the measurement soil resistance R_f under the influence of the crack width W, crack depth D and complex crack conditions. Based on the study above, the quantitative relationship between crack depth and the electrode distances is obtained, which is applied to measure the crack depth of an expansive soil slope and verified.

2 Experimental analysis

2.1 Electrical resistance method

The resistance of soil is usually tested by a four-electrode measurement system. According to Ohm's law, the resistance of soil with a single crack can be expressed as (Figure 1):

$$R_f = \frac{\Delta U}{I} = \frac{1}{K}\rho = \frac{1}{K}\rho[w, T, F(W, D, L)] \quad (1)$$

$$K = 2\pi \Big/ \frac{1}{AM} - \frac{1}{AN} - \frac{1}{BM} + \frac{1}{BN} \quad (2)$$

where I is the injected current (A); ΔU is the electric potential difference between M and N; MA, MB, NA and NB represent the relative spacing (m) between the electrodes M and A, M and B, N and A, and N and B, respectively; K is the electrode coefficient, which represents the geometry and the spatial position of the electrodes and comprehensive influence on the results caused by the electrodes; F is the crack influence function, which is related to the crack width W, crack depth D and crack length L; and ρ is the resistivity function,

TABLE 1 Mineral composition and physical properties of the clay.

Mineral composition	Content (%)	Soil properties	Values
Montmorillonite	12	Free swelling ratio	50.0%
Illite	17	Specific gravity	2.71
Hydromica	18	Plastic limit	25.3%
Chlorite	24	Liquid limit	55.0%
Kaolinite	8	Dry density	1.49 g/cm^3
Quartz	10	USCS classification	CH
Feldspar	2		
Calcite	9		

the value of which is mainly related to the moisture capacity w, the temperature T and the crack influence function F for the same soil sample.

The resistance of soil with a single crack is related to the electrode spatial position and electrical resistivity, the two variables of which are analyzed as follows. The soil resistivity is impacted by basic soil physical properties, which include the mineral composition, arrangement, size and shape of soil particles, pore structure and external environmental factors. The environmental factors are made up of many aspects, such as moisture capacity w and temperature T. To explore the regularity between crack development and the resistivity measurement value and decrease the influence of external factors as much as possible, the parameter logarithm of the resistance ratio N_r is defined, and $N_r = \ln(R_f/R_0)$, where R_f is the resistance value of soil with a single crack under certain external conditions and R_0 is the resistance value of the identical soil sample with no fissure under the same external conditions. In the site measurement process, when the moisture capacity w and temperature T are controlled at certain values, R_0 is a constant value, and the value of N_r is only related to the crack influence function F. During the study, the test electrode pair is almost symmetrically distributed perpendicular to the crack, which means that the spatial position of the electrodes only has a correlation with distances between two electrodes d. The problem is transformed to studying the ascending trend of the N_r value with the development of the crack and under the same crack development condition and the effect of different distances between two electrodes d on the R_f value.

2.2 Material

The mineral composition and physical properties of the expansive soil collected from the Central Line Project of South-to-North Water Diversion are summarized in Table 1. According to the USCS classification, it is a fat clay (CH) (ASTM, 2011). To ensure that there are no impurities in the soil sample that could lead to some disturbances, the coarse particles in the soil sample should be eliminated after retrieval of the soil sample.

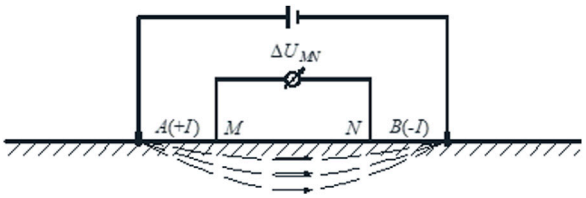


FIGURE 1
Schematic view of the four-electrode method: dashed lines indicate the current flow through the medium (An et al., 2019).

2.3 Specimen preparation

The natural soil retrieved from the field was oven-dried, crushed and sieved in the laboratory. The soil powder was homogeneously mixed with water to reach an initial water content of 20%, and then a soil square mold with an internal size of 400 mm in both the length and width and 200 mm in height was made for specimen preparation in this study. The soil mold was placed in a special Plexiglas mold and then sealed in an air-proof plastic bag for 72 h to allow a uniform distribution of moisture within the soil in the underground room at a temperature of 20 °C. The test electrodes with a length of 17.0 mm and a diameter of 1.5 mm were fabricated by plating platinum black. The part of the electrode embedded in the soil is approximately 2 mm, which is much less than the value of the soil sickness to minimize the influence of the electrode size on the measurement resistance R_f. To reduce the disturbance of the electrode arrangement on the measurement resistance R_f, the current electrodes and the test electrodes are located on the same measurement points. The electrodes were symmetrically arranged perpendicular to the crack development plane.

2.4 Experimental method

The integrated experimental setup used in this study consists of a specially designed sample mold and three monitoring systems, as shown in Figure 2: an electrical resistance acquisition (ERA) system (electrical resistance measuring equipment (EMA) and connective electrodes), a temperature and relative humidity (RH) monitoring (T/RHM) system (TES-1370, with an accuracy of±0.1°C for temperature and ±3% for RH) and a data analysis (DA) system (monitoring system). The details of the equipment are shown in Figure 2. The electrical resistance measuring equipment has the function of power supply, control, transmission, storage and display. Meanwhile, the integrated experimental equipment could be used in the engineering field because of the use of anti-interference means such as configuration of the stabilized filter chip. The T/RHM system was installed close to the soil sample to monitor the variations in ambient temperature and relative humidity (RH) during the test. All data and images were automatically recorded and displayed by a computer connected to these systems. The DA system was employed to postprocess the data and draw the initial data curve.

In the experiments, seven different crack depth D values and five pairs of measuring points were designed to study the variation rules

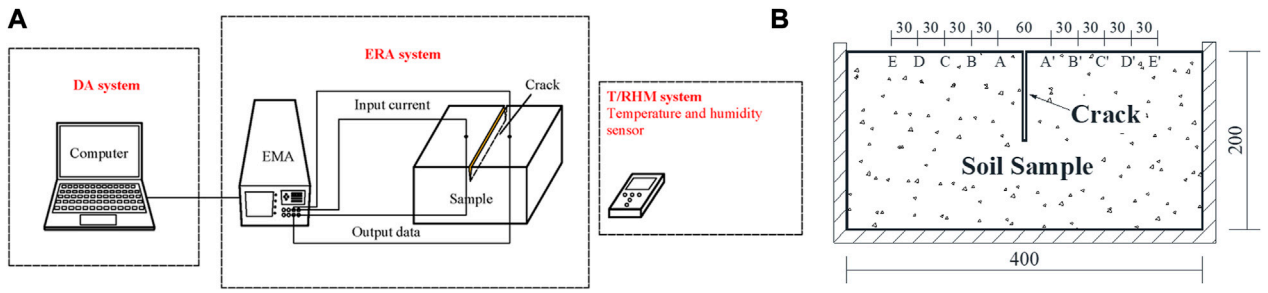


FIGURE 2
(A) Schematic diagram of the experimental setup, including the soil samples, ERA system, T/RHM system and DA system; **(B)** Schematic view of the experimental mold (unit: mm).

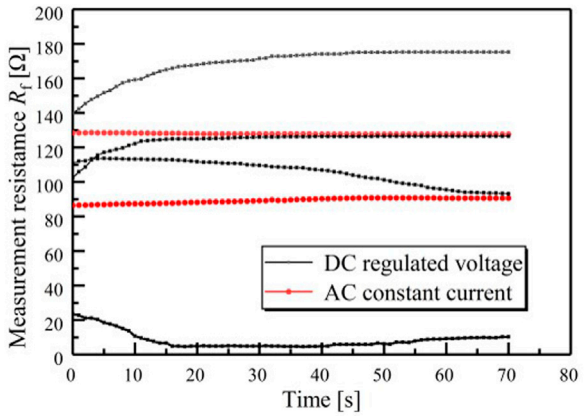


FIGURE 3
Changes in soil resistance R_f values with time.

of the logarithm of the resistance ratio N_r values when the crack developed along the depth direction at each electrode distance d and the influence of electrode distance d change on soil resistance measurement R_f at each crack depth D. During the experimental procedure, the development of the crack was simulated by cutting the crack in the center of the sample with a soil adjusting knife. To restore the whole crack development process as much as possible, the crack development depth D values were selected as 0 cm, 2 cm, 3 cm, 4 cm, 6 cm, 8 cm and 10 cm. By changing the position of the measuring points, the electrode distance d was changed to simulate different electrode placement positions. The electrode distance d values were 6 cm (AA'), 12 cm (BB'), 18 cm (CC'), 24 cm (DD') and 36 cm (EE'). The arrangement of the measuring points is shown in Figure 2B.

The soil particles and water in the soil are both polarizable materials. During the experimental procedure, the resistance measurement file of the multimeter was employed to discover the variation rules of the resistance measurement R_f value of different soils with time under the circumstances of direct current regulated voltage (DC regulated voltage) and alternating current constant current (AC constant current), which is shown in Figure 3. The soil resistance measurement R_f values exhibited irregular changes under DC regulated power supply mode, while the values were relatively stable over time under AC constant current power supply mode. To ensure the repeatability of the experiments, the AC constant current power supply mode was adopted. Although the measurement resistance R_f values tested by this method tended to be influenced by the soil capacitance and inductive impedance, the results were relatively stable and had little effect on the analysis of the measurement resistance R_f under the development of the soil crack.

After selection of the power supply, the test electrode pair was arranged at the AA' measurement point to test the resistance R_f value of the soil with no crack. During the experimental procedure, a soil knife was used to cut a crack running through the soil, the depth of which was 2 cm. Then, the location of the electrode pair was changed from BB testing points to EE testing points to obtain the measurement resistance R_f values under different electrode distances d when the crack depth D was 2 cm. A soil knife was employed to change the crack depth D to test the measurement resistance R_f values under the circumstances of every crack depth D and five groups of electrode distances d.

2.5 Results and discussion

During the experimental procedure, the measurement resistance R_f values of different crack development depths D and different measurement points were recorded by the DA system. The curve of the logarithm of the resistance ratio N_r and crack depth D and the curve of measurement resistance R_f values and electrode distances d are drawn in accordance with data from the experiments. The experimental results are shown in Table 2, and the curves are presented in Figure 4.

It is known from the variation of N_r values with changes of crack depth D (Figure 4A) that the development of the crack results in partial partition of the soil. Due to the high resistance of air, the resistance of partial soil surges and the measurement resistance values present a noticeable change. The resistance of soil with a single crack increases as the crack develops along the depth direction. Moreover, according to the variation amplitude of the curve in Figure 4A, it is demonstrated that the curves of AA' and BB' are in the shape of an S overall, and the variation amplitude of N_r values rises and then slows with crack development along the depth

TABLE 2 Record sheet of measurement resistance R_f values (Ω).

Crack depth measuring point	0 cm	2 cm	3 cm	4 cm	6 cm	8 cm	10 cm
AA' (6 cm)	38.88	41.15	42.23	45.79	50.77	55.96	60.25
BB' (12 cm)	42.88	43.63	44.49	46.55	50.01	54.66	58.73
CC' (18 cm)	45.26	45.80	46.77	48.07	51.09	54.98	58.83
DD' (24 cm)	48.28	49.04	49.90	51.31	53.69	57.58	62.08
EE' (36 cm)	54.44	54.66	55.85	56.82	58.77	62.55	66.70

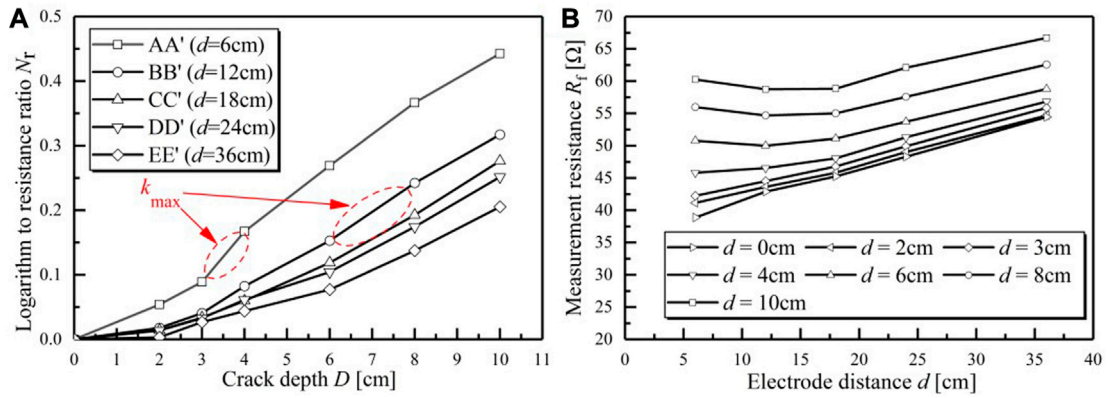


FIGURE 4
(A) Variation in N_r values with changes in crack depth D; **(B)** Variation of measurement resistance R_f values with changes in the electrode distance d.

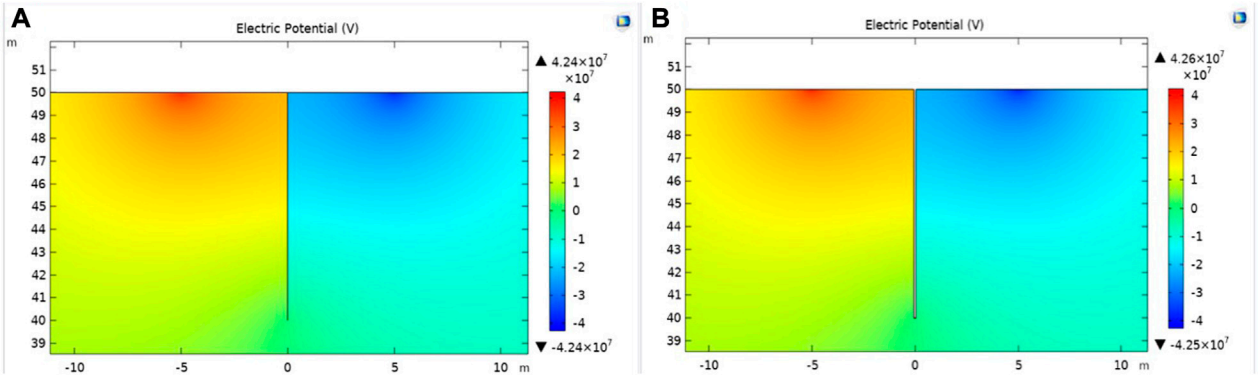


FIGURE 5
Electric potential iso-surface schematic views with changes in the crack width (D=10 m): **(A)** W = 0.001 m; **(B)** W = 0.1 m.

direction. When the crack depth D is approximately half of the electrode distance d, the corresponding variation amplitude of the AA' and BB' curves reaches the maximum. In detail, the corresponding variation amplitude of the AA' and BB' curves reaches the maximum (k_{max}) when the crack depth is approximately 3–4 cm for the AA' curve and the crack is in the depth range of 6–8 cm for the BB' curve. The curves of CC', DD' and EE' show an upward trend, and there is a growing increase in the variation amplitude.

Figure 4B illustrates the variation in the resistance R_f values with changes in the electrode distance d. It is demonstrated that the soil resistance increases constantly as the crack develops along the depth direction. There is a linear relationship between the electrode distance d and the measurement resistance R_f values with no crack. When the crack develops, the curves are in the shape of a U overall. Each curve has the lowest point where each R_{min} value corresponds to a specific d value.

This phenomenon is due to the existence of an inhomogeneous medium, which results in distortion of the electric field. The specific

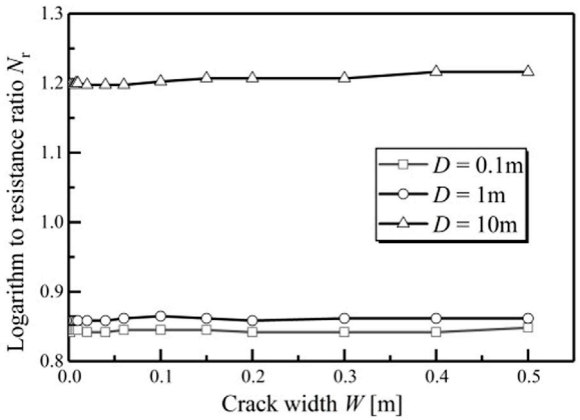


FIGURE 6
N_r-W curve with the change in the crack depth D.

degree of distortion impact depends on the influence range of the electric field. The current between electrode pairs is conducted through the soil. When the crack impedes current conduction between electric pairs, the measurement resistance R_f values of the electric pair decrease. The measurement resistance R_f values of the electric pair would not be affected if there is no block of the crack. It could be speculated that the N_r value depends on the morphological characteristics of the crack. There exists a regular variation in the N_r value with the crack developing along the depth direction. Meanwhile, there is an underlying regulation of the electrode distance d corresponding to the R_{min} value in each R_f-d curve. The quantitative relationship is further explored by numerical simulation.

3 Numerical simulations

3.1 Numerical models and boundary conditions

There are inherent drawbacks for laboratory tests in regard to the influence of the crack width W, crack depth D, crack length L and complex crack conditions on the measurement resistance R_f values. Owing to the limitation of the model sample size and discontinuity of the crack depth D values in laboratory tests, it is difficult to explore the specific crack depth D value when the slope of the N_r-D curve reaches the peak and the specific electrode distance d corresponding to the R_{min} value in the R_f-d curve. To further study the quantitative relation and probe into the possibility of crack depth D inversion, numerical analysis is conducted below.

The influence range of the electric field is spherical and has the obvious characteristic of symmetry. For the primary crack at the site, more than one pair of electrodes is arranged from end to end of the surface crack to portray a general view of the crack underground, owing to the effective range of the electric field. Therefore, the two-dimensional profile with crack width W and crack depth D is taken to conduct further study.

The two-dimensional finite element model of soil with a single crack is established by the software COMSOL Multiphysics, and the interface of static electricity is conducted to perform the numerical calculation.

Because the actual geological conditions of simulated soil with a single crack are complicated, appropriate assumptions are made according to geological conditions to ensure the reliability of numerical analysis: (*a*) The simulated soil area is a semi-infinite space. (*b*) Soil is a continuous homogeneous medium, and the physical parameters of the soil are regarded as constants. The change in the electric field has no effect on the physical properties of the soil. (*c*) In the process of crack development, the soil does not deform in any direction.

The calculation region of the two-dimensional finite element model of soil with a single crack is 200 m × 100 m. The power supply electrodes and the measuring electrodes are arranged at the same measuring point. The pair of electrodes contains a constant current input source and a constant current output source. The surface current density is set to ±0.1 C/m^2. The crack resistivity is 100,000 Ω m, and the soil resistivity is 100 Ω m. The Earth surface is set as an insulation boundary and the ground boundaries along with the infinite ground to the Earth.

The calculation must meet the following conditions.

Constant current equation:

$$\nabla \cdot \boldsymbol{J} + \frac{\partial \nabla \cdot \boldsymbol{D}}{\partial t} = 0 \tag{3}$$

Constitutive relation for linear materials:

$$\boldsymbol{D} = \varepsilon_0 \varepsilon_r \boldsymbol{E} \tag{4}$$

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{5}$$

Electric potential equation:

$$-\nabla V = \boldsymbol{E} \tag{6}$$

Combining Eqs 3 and 4, 5 and 6, the following can be obtained:

$$-\nabla \cdot \left(\sigma \nabla + \frac{\partial (\varepsilon_0 \varepsilon_r \nabla V)}{\partial t} \right) = 0 \tag{7}$$

Electrostatic quasi-static equation:

$$\boldsymbol{n} \cdot (\boldsymbol{J}_2 - \boldsymbol{J}_1) = \boldsymbol{J}_n \tag{8}$$

$$\boldsymbol{n} \cdot (\boldsymbol{E}_2 - \boldsymbol{E}_1) = 0 \tag{9}$$

In the equations, ∇ is the Hamiltonian operator; $\boldsymbol{J}$ is the current density vector; $\boldsymbol{D}$ is the electric displacement field; t is time; ε_0 represents the dielectric constant in free space; ε_r represents the relative dielectric constant; $\boldsymbol{E}$ is the electric field intensity; σ is electrical conductivity; V is the electric potential; and $\boldsymbol{J}_n$ is the current along the normal phase.

The regular patterns of influence on the electric field of soil containing a single crack considering the crack width W, crack depth D and the complex crack conditions are discussed as follows. The study is carried out from qualitative and quantitative angles, i.e., qualitatively observing the electric field distribution under the development of the crack and quantitatively exploring the change in the N_r value under the development of the crack and the R_f value with different electrode distances d under the same crack depth.

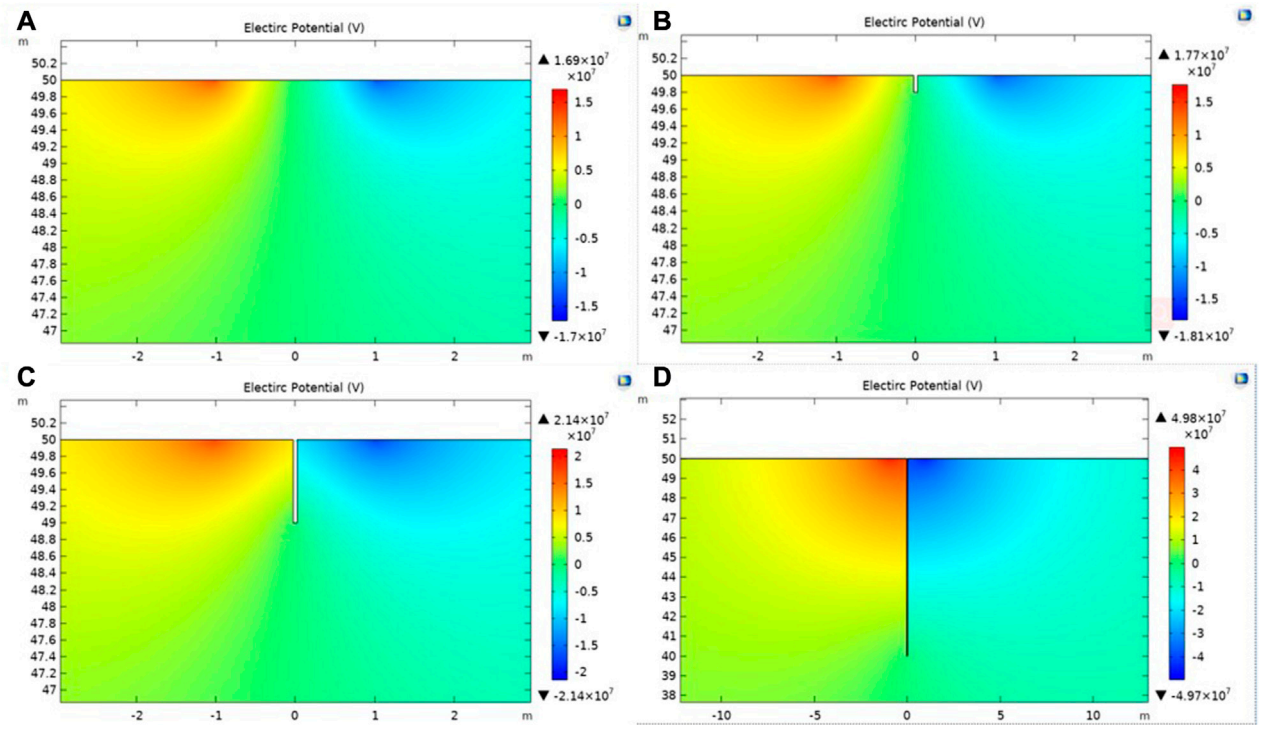


FIGURE 7
Electric potential iso-surface schematic views with changes in the crack depth D (d=2 m): **(A)** D= 0 m; **(B)** D= 0.2 m; **(C)** D = 2 m; **(D)** D = 10 m.

3.2 Crack width W

Considering the practical crack conditions, the study explores the changes in N_r values when the crack width W ranges from 0.001 m to 0.5 m at different crack depth D values. The profile of the geoelectric model of soil with a single crack is obtained to carry out analysis of the numerical simulation results under the change in W.

The equipotential surface schematic views of planforms with the change in the crack width W are shown in Figure 5. According to the electric field distribution diagrams (Figure 5), the change in the crack width W has no obvious influence on the electric field distribution when the W value changes from 0.001 m to 0.1 m.

Taking crack depths of 0.1 m, 1 m, and 10 m as examples, the soil electrical parameters could be adopted with the change in the crack width W ranging from 0.001 m to 0.5. Based on the soil electrical parameters, the quantitative relationship of the logarithm of the resistance ratio N_r and the crack width W will be explored.

Figure 6 shows the N_r-W curve with the change in the crack depth D. The N_r value has a remarkable increase with the crack developing along the crack depth direction. Compared to the change in the N_r value caused by the crack developing along the depth direction, the N_r value barely changes with the crack developing along the width direction. Based on the basic theory of electric fields, it is speculated that the trend of the crack width is consistent with the trend of the electric field lines, which are unable to impede the conduction of current between two electrodes. In the following calculation and analysis, the crack width W is set to 0.05 m.

3.3 Crack depth D

Keeping the crack width constant at W=0.05 m, the profiles of soil with a single crack are obtained to conduct an analysis of two-dimensional numerical simulation with the change in the crack depth D.

Figure 7 demonstrates the equipotential surface schematic views with changes in the crack depth D. The electric field distribution diagrams above illustrate the regulations that with regard to the semi-infinite uniform layer, current is transmitted homogeneously through the soil when the crack is not developed (a). When the crack develops from the ground along the crack depth direction and the depth of the crack is shallow, the distribution of the electric field in the soil changes slightly (b). When the crack depth D exceeds the electrode distance d, the electric field distribution generates obvious distortion, and the field presents an overt distribution along the crack depth direction at the crack location, owing to the impediment of current transfer by the crack (c and d).

From a quantitative perspective, the study is divided into two aspects: the ascending trend of the N_r value with the development of the crack and the effect of different electrode distances d on the R_f value at the same crack depth D. In regard to the study of N_r values, the electrode distances d are set to constant values. To facilitate the analysis and comparison of data, the study defines a parameter crack depth ratio D_r, which means the crack depth D is divided by the electrode distance d. In the process of numerical simulation, the test electrode pair is symmetrically distributed almost perpendicular to the crack. The electrode distance d values are set to 1 m, 2 m, 3 m, 4 m, and 5 m. The D_r values range from 0 to 10, and the crack width

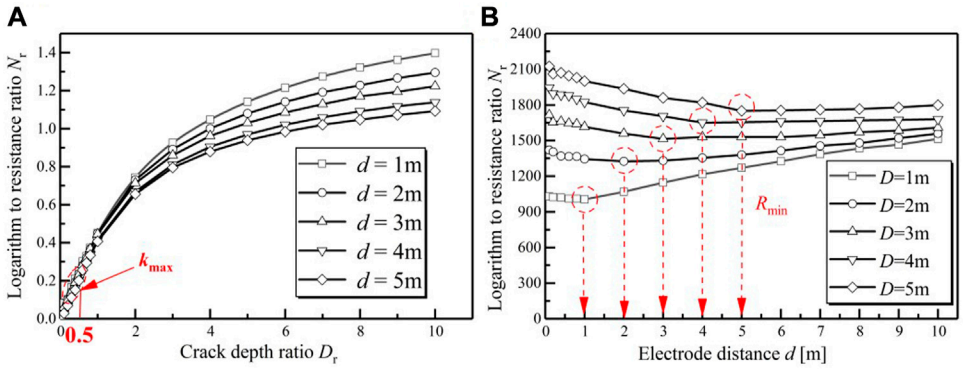


FIGURE 8
(A) N_r- D_r curve with changes in the electrode distance d; **(B)** R_f-d curve with changes in the crack depth D.

W is set to 0.05 m. The N_r-D_r curve is drawn to explore the dynamic variation regulations.

The crack width W value is set to 0.05 m, and the crack depth D values are set to 1 m, 2 m, 3 m, 4 m, and 5 m. The test electrode pair is symmetrically distributed almost perpendicular to the crack, and the electrode distance d values range from 0.1 m to 10 m. The curve of R_f-d is drawn to explore the quantitative relation of the parameters D and d.

It is demonstrated that the N_r-D_r curve of the crack is in the shape of an S overall (Figure 8). The slope of the curve is defined as k, and the value of k shows an increasing and decreasing trend. The change in the curve manifests obvious periodic characteristics. In the first period, when the depth of the developing crack is shallow, the increase in the amplitude of the N_r value is relatively slow, and the k value is small. In the second period, the increase amplitude of the N_r value is great. There exists an explosion for the k value, which reaches the peak when D_r is approximately equal to 0.5. In the third period, when D_r surpasses 1, the N_r value does not grow substantially and only increases slightly. The k value begins to wane.

Figure 8 demonstrates that the R_f-d curve is in the shape of a U. The value of the resistance value of the soil with the single crack R_f starts from the maximum value and gradually decreases to the minimum value R_{min}. Subsequently, the R_f value presents the tendency of a moderate increase. Thus, for estimating the unknown crack depth D, the operators could change the value d in the vicinity of the crack to measure the relative change in the R_f value. The variation curve R_f-d is drawn to determine the value d corresponding to the R_{min} value. At this point, the crack depth D is approximately equal to the electrode distance d, that is, $D \approx d$.

The findings could be drawn as a new method, i.e., the electrical resistance method, which could be employed to conduct crack depth measurements in engineering projects. Regarding the primary crack, electrode pairs are arranged from end to end of the surface crack in accordance with the measurement points. R_f-d curves could be drawn by changing the electrode distance d values. The crack depth D value in each measurement point pair is approximately equal to the electrode distance d corresponding to R_{min}. Combining every crack depth D value could obtain the general view of the underground crack by this method. However, the method requires the electrode pair to be symmetrically distributed perpendicular to the crack. which is difficult to realize *in situ*. Therefore, the influence of asymmetry measurements on the results is studied below.

3.4 Complex crack conditions

Symmetry electrodes and the ideal crack location are quite hard to realize in actual measurements, so it is of great necessity to analyze the influence of asymmetry measurements on the results of the soil resistance test and the crack depth measurement method. The profile of a two-dimensional finite element model of soil with a single crack is carried out to conduct the study. The study considers the conditions, including the relocation, the offset and the rotation of the crack, and changes in the relative position to explore the electric field change rule of soil with a single crack model (Figures 9, 12, 13).

The crack depth D is set to 1 m, the crack width W is set to 0.05 m and the electrode distance d is set to 1 m. To represent the relative position relationship of the crack, the parameters of the relocation distance ratio of the crack W_r, the offset distance ratio of the crack P_r and the rotation angle of the crack θ are defined. The W_r value stands for the ratio of the crack outward displacements along the crack length direction to the crack depth D. The P_r value means the ratio of the crack offset displacements between the electrode pair to the electrode distance d. When deliberating θ, it is stipulated that the rotation angle of the crack perpendicular to the electrode line is 0° and $0° \leq \theta \leq 90°$.

The W_r value ranges from 0 to 5, which means that the crack outward displacements along the crack length direction range from 0 m to 5 m. The N_r-W_r curve is drawn in the process of crack outshift. The P_r value ranges from −0.4 to 0.4, which means that the crack offset displacements between the electrode pair range from −0.4 m to 0.4 m. The N_r-P_r curve is drawn in the process of crack offset. In the study process, the rotation angle of the crack perpendicular to the electrode line is 0°, and the θ value ranges from 0° to 90°. The N_r-θ curve is drawn in the process of crack rotation. To make a comparison, one N_r-D_r curve is set as the standard situation.

In terms of the N_r-W_r curve (Figure 10), the N_r value continues to decrease as W_r changes from 0 to 5. When W_r is equal to 0.5, the N_r value is half of that of the crack in the middle of the electrodes (the standard situation in Figure 10). According to the N_r-P_r curve

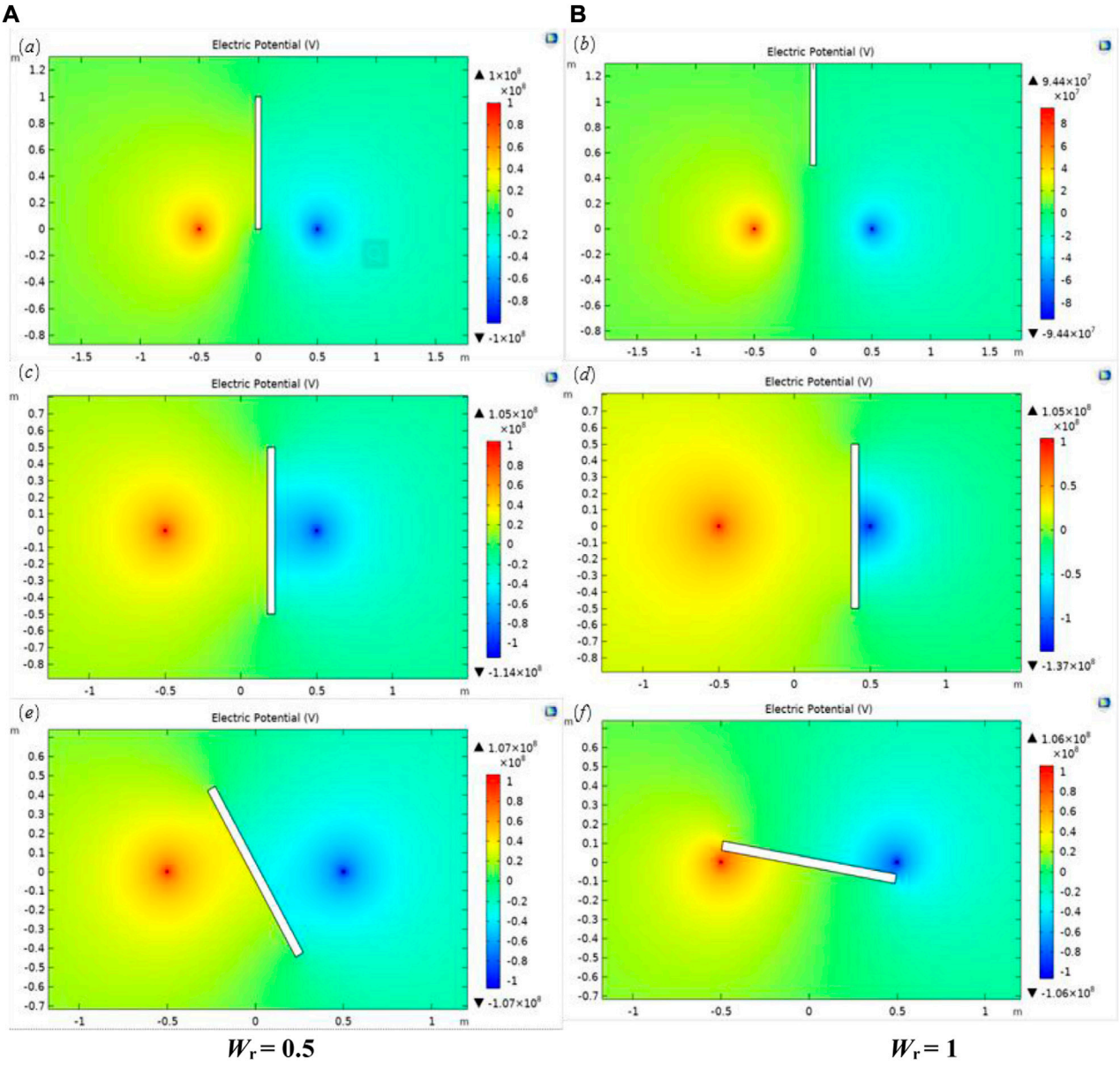


FIGURE 9
Electric potential iso-surface schematic view with changes in the relocation, offset and rotation of the crack: **(A)** P_r = 0.2; **(B)** P_r = 0.4; **(C)** W_r = 0.5; **(D)** W_r = 1; **(E)** θ = 30°; **(F)** θ = 80°.

(Figure 10), the curve presents a downward trend, but compared to the increase in the N_r value caused by crack development along the crack depth, the curve declines slowly in Figure 10. In other words, in the actual measurement process, there is some asymmetry when the electrodes are embedded, which has little influence on the measurement results. In light of the N_r-θ curve, the curve shows an upward trend, but compared to the increase in the N_r value caused by crack development along the crack depth (the standard situation in Figure 10), the curve rises slowly in Figure 10. In particular, when θ is less than or equal to 40°, the N_r value barely changes.

In the actual measurement process, the electrodes have some asymmetry with the crack when buried, which has little influence on the measurement results. When the crack exists within the influence range of the electric field, it could impede current conduction between electrodes, but the measurement results have little correlation with the specific position of the crack in the electric field.

4 Practical engineering

4.1 Engineering situation

An expansive soil slope in Jiangsu Province in China is typical of slope instability in landscape areas. The landslide still creeped when it rained. Obvious landslide features such as crack phenomenon,

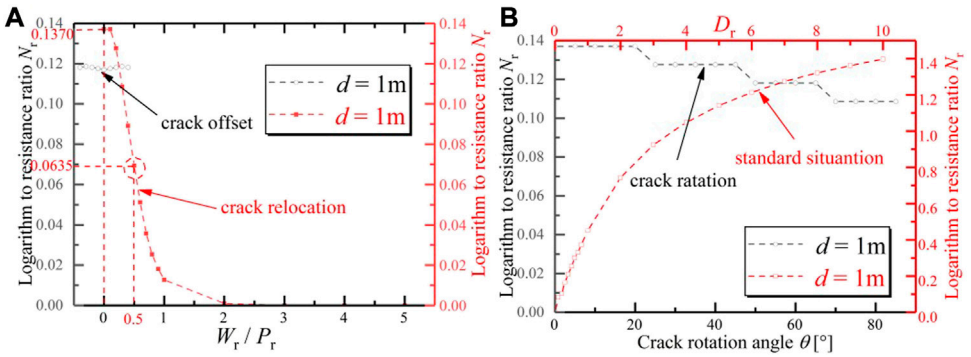


FIGURE 10
(A) N_r-W_r curve and N_r-P_r; **(B)** N_r-θ curve.

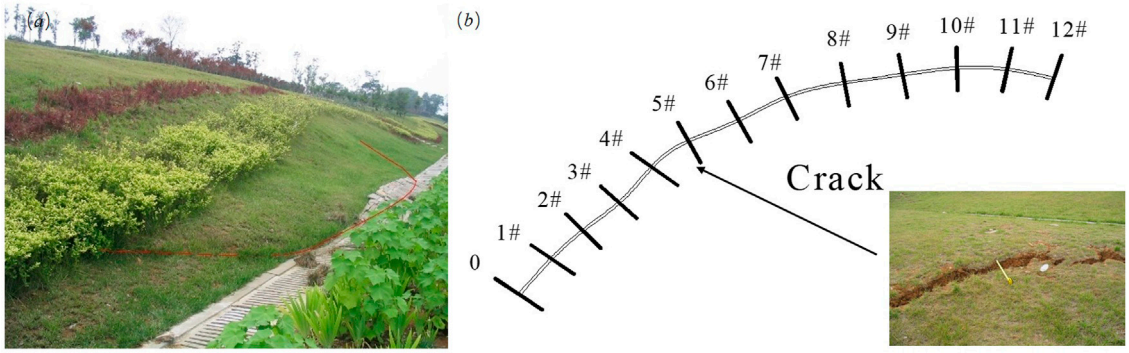


FIGURE 11
(A) Expansive slope landslide; **(B)** Survey profile diagram.

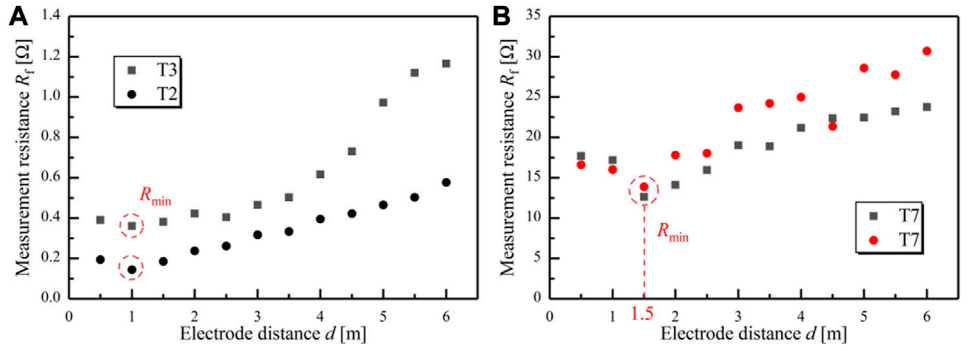


FIGURE 12
(A) Diagram of measurement resistance R_f after rainfall; **(B)** Measurement resistance R_f after a week without rain.

collapse, and hump emerge in the landscape area. Different types of cracks appear in the upper part, both sides and the lower part of the landslide. Under the influence of rain, the landslide continued from time to time, posing a serious threat to traffic and life safety. The landslide has a maximum elevation of 27.0 m at the back edge and 10.0 m at the front edge with a slope of 11°. Generally, the slope is relatively slow.

The physical and geological phenomena of landslides are relatively well developed, including cracks and collapses. The width of the cracks is between 5 cm and 15 cm. The depth is

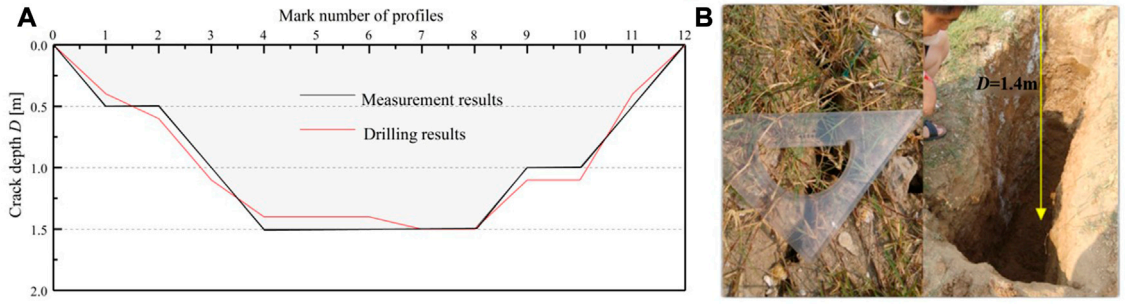


FIGURE 13
(A) Crack depth distribution map; **(B)** Drilling result of #6.

approximately 2.0 m, and shallow cracks are approximately 0.5 m. The wider cracks are mostly oriented in the east–west direction and mostly arcuate. There are three main cracks on both sides of the landslide body, which are often accompanied by small feather-like cracks parallel to the sliding direction. The lower part of the landslide is obviously swollen, forming a typical slide tongue of the landslide. The main crack of the landslide is taken as an example to conduct crack depth measurements by means of the soil resistance method.

4.2 Measurement procedure

For the measurement of the depth of the main crack in expansive soil, 13 measurement sections, i.e., #0–12, were selected along the vertical direction of the crack (Figure 11). The sections were not parallel and were approximately 1 m apart. A metal rod was used as the measuring electrode. The metal rod was inserted into the soil up to 10 cm to ensure good contact with the soil and no shaking of the metal rod. The electrode layout was perpendicular to the fracture profile and symmetrically distributed along the crack. The electrode distance d ranged from 1 m to 11 m. Data were recorded and analyzed by the DA system. It must be guaranteed that there was no rain during the measurement procedure, which began 1 week after rain. During the process of water absorption and loss of the expansive soil slope, the expansion and contraction cracks developed and closed along the depth direction.

4.3 Results

The central profile (#6) was selected for analysis, and the curves of resistance measurement R_f and electrode distance d on the second day after rain, the third day after rain and 1 week after rain were drawn, which were denoted T2, T3 and T7, respectively. The resistance 1 week after rain was measured repeatedly (Figure 12).

The data showed (Figure 12) that as time progressed, the soil cracks developed along the depth direction due to the drying shrinkage, resulting in an increase in the resistance measured between electrodes. The electrode distance d values corresponding to the lowest point of the T1, T2 and T7 curves are approximately 1 m, 1 m and 1.5 m, respectively. According to the above rules of crack depth D and electrode distance d, the crack depths of T1, T2 and T7 are 1 m, 1 m and 1.5 m, respectively. Similarly, 11 profiles were measured on the same day to detect the development depth of fractures at different profiles, and a schematic diagram of the depth and shape of fractures in the longitudinal profile was drawn (Figure 13). To verify the probe results, a test pit survey was conducted, and the results are shown in Figure 13. The crack depth measured by the electric resistance method is consistent with the drilling results.

5 Conclusion

This study explores a new method to conduct crack depth measurements in engineering projects by means of combining laboratory tests and numerical modeling analysis. The following conclusions can be drawn:

(1) The laboratory tests determined that there exists a regular variation in the N_r value with the crack developing along the depth direction. Meanwhile, there is an underlying regulation of the electrode distance d corresponding to the R_{min} value in each R_f-d curve.
(2) The numerical modeling analysis further shows that the N_r-D_r curve is in the shape of an S. The slope of the curve reaches the peak when D_r is approximately equal to 0.5. The R_f-d curve is in the shape of a U. The d value corresponding to the minimum value R_{min} is approximately equal to the crack depth D. The N_r value barely changes with the change in crack width. In the range of the electric field, the offset and rotation of the crack have little effect, while the measurement results have a strong dependence on relocation movement.
(3) Regarding the primary crack, electrode pairs are arranged from end to end of the surface crack in accordance with the measurement points. R_f-d curves could be drawn by changing the electrode distance d values. The crack depth D value in each measurement point pair is approximately equal to the electrode distance d corresponding to R_{min}. By means of this electrical resistance method, a general view of the crack underground could be portrayed. This method was employed to measure the crack depth of an expansive soil slope and

verified by comparing the measurement results and drilling results.

In this paper, the quantitative relationship between crack development and electrical parameter changes is confirmed from the point of view of experiments and numerical simulations. However, it should be pointed out that the development and evolution of soil cracks exist in three-dimensional space, so it is inevitable that the geometric parameters of cracks cannot be accurately measured in two-dimensional studies, and there is some deviation from the actual situation. However, for the prevention and control of expansive soil slope engineering, the deviation is within the allowable range. In the future, the application of the electric resistance method in the development of three-dimensional soil cracks will be further explored.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.Conceptualization, JG and LC; methodology, JG, YW and SY; validation, LC; formal analysis, LH; investigation, JG, LH, YW, and YT; resources, LC; data curation, JW and HL; writing-original draft preparation, JG and YW; writing-review and editing, JG; funding acquisition, LC.

Funding

This work was supported by the National Key Research and Development Program (No. 2019YFC1510802), the National Natural Science Foundation of China (Grant No. U2243244), the major science and technology project of the Ministry of Water Resources (SKR-2022072).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ackerson, J. P., Morgan, C. L. S., Everett, M. E., and McInnes, K. J. (2014). The role of water content in electrical resistivity tomography of a vertisol. *Soil Sci. Soc. Am. J.* 78, 1552–1562. doi:10.2136/sssaj2014.01.0032

Albrecht, B. A., and Benson, C. H. (2001). Effect of desiccation on compacted natural clays. *J. Geotechnical Geoenvironmental Eng.* 128 (1), 67–75. doi:10.1061/(asce)1090-0241(2001)127:1(67)

An, N., Tang, C. S., Cheng, Q., Wang, D. Y., and Shi, B. (2019). Application of electrical resistivity method in the characterization of 2d desiccation cracking process of clayey soil. *Eng. Geol.* 265, 105416. doi:10.1016/j.enggeo.2019.105416

Baer, J. U., Kent, T. F., and Anderson, S. H. (2009). Image analysis and fractal geometry to characterize soil desiccation cracks. *Geoderma* 154 (1-2), 153–163. doi:10.1016/j.geoderma.2009.10.008

Baram, S., Ronen, Z., Kurtzman, D., Külls, C., and Dahan, O. (2013). Desiccation-crack-induced salinization in deep clay sediment. *Hydrology Earth Syst. Sci.* 17 (4), 1533–1545. doi:10.5194/hess-17-1533-2013

Cai, Y. J., and Li, L. (2018). "Engineering characteristics of expansive soil in the middle line of south-to-north Water Transfer project and failure mechanism of slope sliding," in *Proceedings of 2018 national annual conference on engineering geology*.

Cała, M., Cyran, K., Stopkowicz, A., Kolano, M., and Szczygielski, M. (2016). Preliminary application of X-ray computed tomograph on characterisation of polish gas shale mechanical properties. *Rock Mech. Rock Eng.* 49, 4935–4943. doi:10.1007/s00603-016-1045-6

Chaduvula, U., Viswanadham, B. V. S., and Kodikara, J. (2017). A study on desiccation cracking behavior of polyester fiber-reinforced expansive clay. *Appl. Clay Sci.* 142, 163–172. doi:10.1016/j.clay.2017.02.008

Chertkov, V. Y. (2000). Using surface crack spacing to predict crack network geometry in swelling soils. *Soil Sci. Soc. Am. J.* 64 (6), 1918–1921. doi:10.2136/sssaj2000.6461918x

Costa, S., Kodikara, J., and Shannon, B. (2013). Salient factors controlling desiccation cracking of clay in laboratory experiments. *Geotechnique* 63 (1), 18–29. doi:10.1680/geot.9.P.105

Dai, A. (2011). Drought under global warming: A review. *Wiley Interdiscip. Rev. Clim. Change* 2, 45–65. doi:10.1002/wcc.81

Damavandi, K., Abedi, M., Norouzi, G. H., and Mojarab, M. (2022). Geoelectrical modelling of a landslide surface through an unstructured mesh. *Bollettino di Geofisica Teorica ed Applicata* 63 (2), 337–356. doi:10.4430/bgo00384

Dyer, M. (2005). Performance of flood embankments in England and wales. *Water Manag.* 157 (4), 177–186. doi:10.1680/wama.2004.157.4.177

Dyer, M., Utili, S., and Zielinski, M. (2009). Field survey of desiccation fissuring of flood embankments. *Water Manag.* 162 (3), 221–232. doi:10.1680/wama.2009.162.3.221

Fma, A., Meng, H., Mn, C., Ka, C., Ay, D., and Aa, E. (2022). Utilizing nwcr optimized arrays for 2d ert survey to identify subsurface structures at penang island, Malaysia. *J. Appl. Geophys.* 196, 104518. doi:10.1016/j.jappgeo.2021.104518

Gunn, D. A., Chambers, J. E., Uhlemann, S., Wilkinson, P. B., Meldrum, P. I., Dijkstra, T. A., et al. (2015). Moisture monitoring in clay embankments using electrical resistivity tomography. *Constr. Build. Mat.* 92, 82–94. doi:10.1016/j.conbuildmat.2014.06.007

Hallett, P. D., and Newson, T. A. (2011). Desiccation of a sensitive clay: Application of the model crack: Discussion. *Can. Geotechnical J.* 35 (6), 1109–1110. doi:10.1139/t99-003

Hewitt, P. J., and Philip, L. K. (1999). Problems of clay desiccation in composite lining systems. *Eng. Geol.* 53 (2), 107–113. doi:10.1016/S0013-7952(99)00023-X

Hirobe, S., and Oguni, K. (2016). Coupling analysis of pattern formation in desiccation cracks. *Comput. Methods Appl. Mech. Eng.* 307, 470–488. doi:10.1016/j.cma.2016.04.029

Jones, G., Zielinski, M., and Sentenac, P. (2012). Mapping desiccation fissures using 3-d electrical resistivity tomography. *J. Appl. Geophys.* 84, 39–51. doi:10.1016/j.jappgeo.2012.06.002

Jones, G., Sentenac, P., and Zielinski, M. (2014). Desiccation cracking detection using 2-D and 3-D electrical resistivity tomography: Validation on a flood embankment. *J. Appl. Geophys.* 106, 196–211. doi:10.1016/j.jappgeo.2014.04.018

Khandelwal, S., Sanchez, M., and Medina-Cetina, Z. (2013). Effect of depth of desiccation cracks on earth embankments. *Geo-Congress*, 800–803. doi:10.1061/9780784412787.080

Kim, K. Y., Zhuang, L., Yang, H., Kim, H., and Min, K. B. (2016). Strength anisotropy of berea sandstone: Results of x-ray computed tomography, compression tests, and discrete modeling. *Rock Mech. Rock Eng.* 49 (4), 1201–1210. doi:10.1007/s00603-015-0820-0

Kodikara, J., and Costa, S. (2013). *Desiccation cracking in clayey soils: Mechanisms and modelling*. Springer Berlin Heidelberg, 21–32. doi:10.1007/978-3-642-32492-5_2

Kong, L. W., Wang, M., Guo, A. G., and Wang, Y. (2017). Effect of drying environment on engineering properties of an expansive soil and its microstructure. *J. Mt. Sci.* 14 (06), 1194–1201. doi:10.1007/s11629-017-4430-z

Kong, L. W., Zeng, Z. X., Bai, W., and Wang, M. (2018). Engineering geological properties of weathered swelling mudstones and their effects on the landslides occurrence in the yanji section of the jilin-hunchun high-speed railway. *Bull. Eng. Geol. Environ.* 77 (77), 1491–1503. doi:10.1007/s10064-017-1096-2

Krisnanto, S., Rahardjo, H., Fredlund, D. G., and Leong, E. C. (2014). Mapping of cracked soils and lateral water flow characteristics through a network of cracks. *Eng. Geol.* 172, 12–25. doi:10.1016/j.enggeo.2014.01.002

Lakshmikantha, M. R., Prat, P. C., and Ledesma, A. (2012). Experimental evidence of size effect in soil cracking. *Can. Geotechnical J.* 49 (3), 264–284. doi:10.1139/t11-102

Li, L., Larsen, I., and Holt, R. M. (2015). Laboratory observation and micromechanics-based modelling of sandstone on different scales. *Rock Mech. Rock Eng.* 48 (4), 1407–1422. doi:10.1007/s00603-014-0642-5

Li, J. H., Lu, Z., Guo, L. B., and Zhang, L. M. (2017). Experimental study on soil-water characteristic curve for silty clay with desiccation cracks. *Eng. Geol.* 218, 70–76. doi:10.1016/j.enggeo.2017.01.004

Li, H. D., Tang, C. S., Cheng, Q., Li, S. J., Gong, X. P., and Shi, B. (2019). Tensile strength of clayey soil and the strain analysis based on image processing techniques. *Eng. Geol.* 253, 137–148. doi:10.1016/j.enggeo.2019.03.017

Lozada, C., Caicedo, B., and Thorel, L. (2015). Effects of cracks and desiccation on the bearing capacity of soil deposits. *Géotechnique Lett.* 5 (3), 112–117. doi:10.1680/jgele.15.00021

Lu, Z. H., Chen, Z. H., and Pu, Y. B. (2002). A CT study on the cracks evolution. *Rock Soil Mech.* 23 (4), 417–422. doi:10.3969/j.issn.1000-7598.2002.04.005

Morris, P. H., Graham, J., and Williams, D. J. (1992). Cracking in drying soils. *Can. Geotechnical J.* 29 (2), 263–277. doi:10.1139/t92-030

Neyamadpour, A. (2018). Detection of subsurface cracking depth using electrical resistivity tomography: A case study in masjed-soleiman, Iran. *Constr. Build. Mater.* 191, 1103–1108. doi:10.1016/j.conbuildmat.2018.10.027

Park, Y., You, Y. J., and Kim, H. J. (2009). Electrical resistivity surveys for gold-bearing veins in the yongjang mine, Korea. *J. Geophys. Eng.* 6 (1), 73–81. doi:10.1088/1742-2132/6/1/008

Pidlisecky, A., Haber, E., and Knight, R. (2007). Resinvm3d: A 3d resistivity inversion package. *Geophysics* 72 (2), 1–10. doi:10.1190/1.2402499

Pouya, A., Vo, T. D., Hemmati, S., and Tang, A. M. (2018). Modeling soil desiccation cracking by analytical and numerical approaches. *Int. J. Numer. Anal. Methods Geomechanics* 43 (3), 738–763. doi:10.1002/nag.2887

Ramana, K. V. (1993). Expansive soils: Problems and practice in foundation and pavement engineering. *Eng. Geol.* 35, 136–138. doi:10.1016/0013-7952(93)90076-o

Rayhani, M. H., Yanful, E. K., and Fakher, A. (2007). Desiccation-induced cracking and its effect on the hydraulic conductivity of clayey soils from Iran. *Can. Geotechnical J.* 44 (3), 276–283. doi:10.1139/t06-125

Ruan, L., Luo, R., Wang, B., and Yu, X. H. (2021). Morphological characteristics of crack branching in asphalt mixtures under compression. *Eng. Fract. Mech.* 253, 107884. doi:10.1016/j.engfracmech.2021.107884

Samouëlian, A., Cousin, I., Richard, G., Tabbagh, A., and Bruand, A. (2003). Electrical resistivity imaging for detecting soil cracking at the centimetric scale. *Soil Sci. Soc. Am. J.* 67 (5), 1319–1326. doi:10.2136/sssaj2003.1319

Samouëlian, A., Richard, G., Cousin, I., Guérin, R., Bruand, A., and Tabbagh, A. (2004). Three-dimensional crack monitoring by electrical resistivity measurement. *Eur. J. Soil Sci.* 55, 751–762. doi:10.1111/j.1365-2389.2004.00632.x

Shin, H., and Santamarina, J. C. (2011). Desiccation cracks in saturated fine-grained soils: Particle-level phenomena and effective-stress analysis. *Géotechnique* 61 (11), 961–972. doi:10.1680/geot.8.p.012

Stewart, R. D., Majdi, R., Najm, A., David, E., and Selker, J. S. (2012). Measurement tool for dynamics of soil cracks. *Vadose Zone J.* 11 (2). doi:10.2136/vzj2011.0048

Sultan, S. A., Santos, F. A. M., and Helal, A. (2006). A study of the groundwater seepage at hibis temple using geoelectrical data, kharga oasis, Egypt. *Near Surf. Geophys.* 4 (6), 347–354. doi:10.3997/1873-0604.2006008

Sungkono, S., Husein, A., Prasetyo, H., Bahri, A. S., and Santosa, B. J. (2014). The vlf-em imaging of potential collapse on the lusi embankment. *J. Appl. Geophys.* 109, 218–232. doi:10.1016/j.jappgeo.2014.08.004

Tabbagh, J., Samouëlian, A., Tabbagh, A., and Cousin, I. (2007). Numerical modelling of direct current electrical resistivity for the characterisation of cracks in soils. *J. Appl. Geophys.* 62 (4), 313–323. doi:10.1016/j.jappgeo.2007.01.004

Taghavi, M., Dovoudi, M. H., Amiri-Tokaldany, E., and Darby, S. E. (2010). An analytical method to estimate failure plane angle and tension crack depth for use in riverbank stability analyses. *Geomorphology* 123 (1-2), 74–83. doi:10.1016/j.geomorph.2010.06.017

Tang, C. S., Wang, D. Y., Zhu, C., Zhou, Q. Y., Xu, S. K., and Shi, B. (2018). Characterizing drying-induced clayey soil desiccation cracking process using electrical resistivity method. *Appl. Clay Sci.* 152, 101–112. doi:10.1016/j.clay.2017.11.001

Tollenaar, R. N., Paassen, L., and Jommi, C. (2017). Observations on the desiccation and cracking of clay layers. *Eng. Geol.* 230, 23–31. doi:10.1016/j.enggeo.2017.08.022

Trabelsi, H., Hadrich, B., and Guiras, H. (2018). Evaporation, shrinkage and intrinsic permeability of unsaturated clayey soil: Analytical modelling versus experimental data. *Arabian J. Geosciences* 11 (8), 184. doi:10.1007/s12517-018-3507-5

Vo, T. D., Pouya, A., Hemmati, S., and Tang, A. M. (2017). Numerical modelling of desiccation cracking of clayey soil using a cohesive fracture method. *Comput. Geotechnics* 85, 15–27. doi:10.1016/j.compgeo.2016.12.010

Vo, T. D., Pouya, A., Hemmati, S., and Tang, A. M. (2018). Modelling desiccation crack geometry evolution in clayey soils by analytical and numerical approaches. *Can. Geotechnical J.* 56 (5), 720–729. doi:10.1139/cgj-2018-0105

Wang, C., Zhang, Z. Y., Liu, Y., and Fan, S. M. (2017). Geometric and fractal analysis of dynamic cracking patterns subjected to wetting-drying cycles. *Soil Tillage Res.* 170, 1–13. doi:10.1016/j.still.2017.02.005

Wang, L. L., Tang, C. S., Shi, B., Cui, Y. J., Zhang, G. Q., and Hilary, I. (2018). Nucleation and propagation mechanisms of soil desiccation cracks. *Eng. Geol.* 238, 27–35. doi:10.1016/j.enggeo.2018.03.004

Xie, C., Ni, P., Xu, M., Mei, G., and Zhao, Y. (2020). Combined measure of geometry optimization and vegetation for expansive soil slopes. *Comput. Geotechnics* 123, 103588. doi:10.1016/j.compgeo.2020.103588

Yang, S. Q., Ranjith, P. G., and Gui, Y. L. (2015). Experimental study of mechanical behavior and X-ray micro CT observations of sandstone under conventional triaxial compression. *Geotechnical Test. J.* 38 (2), 179–197. doi:10.1520/GTJ20140209

Yin, Z. Z., Yuan, J. P., Wei, J., Cao, X. S., Liu, H. Q., and Xu, B. (2012). Study on the effects of cracks to the stability of expansive soil slope. *Chin. J. Geotechnical Eng.* 34 (012), 2155–2161. doi:10.3969/j.issn.1008-3383.2014.03.012

Yu, Q. L., Yang, S. Q., Ranjith, P. G., Zhu, W. C., and Yang, T. H. (2016). Numerical modeling of jointed rock under compressive loading using x-ray computerized tomography. *Rock Mech. Rock Eng.* 49 (3), 877–891. doi:10.1007/s00603-015-0800-4

Yuan, J. P., and Yin, Z. Z. (2004). Quantitative index of fissure and strength characteristics of fissured expansive soils. *J. Hydraulic Eng.* (6), 108–113. doi:10.3321/j.issn:0559-9350.2004.06.019

Zhang, J. M., Luo, Y., Zhou, Z., Chong, L., Victor, C., and Zhang, Y. F. (2021). Effects of preferential flow induced by desiccation cracks on slope stability. *Eng. Geol.* 288, 106164. doi:10.1016/j.enggeo.2021.106164

Zhao, Z., and Zhou, X. P. (2020). Establishment of numerical cracking constitutive models using 3d reconstruction and x-ray ct images of geomaterials. *Int. J. Mech. Sci.* 183 (22), 105814. doi:10.1016/j.ijmecsci.2020.105814

Zhen, S. H., Jin, J. L., Yao, H. L., and Ge, X. R. (2006). Analysis of initial cracking behavior of expansive soil due to evaporation. *Rock Soil Mech.* 27 (12), 2229–2233. doi:10.3969/j.issn.1000-7598.2006.12.027

Zhu, L., Shen, T., Ma, R., Fan, D., and Zha, Y. (2020). Development of cracks in soil: An improved physical model. *Geoderma* 1 (1), 114258. doi:10.1016/j.geoderma.2020.114258

frontiers | Frontiers in Earth Science

TYPE Methods
PUBLISHED 22 March 2023
DOI 10.3389/feart.2023.1127889

OPEN ACCESS

EDITED BY
Wei Zhao,
Institute of Mountain Hazards and Environment (CAS), China

REVIEWED BY
Yi Wang,
China University of Geosciences Wuhan, China
Haijia Wen,
Chongqing University, China

*CORRESPONDENCE
Tianqiang Qu,
2149859375@qq.com

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 20 December 2022
ACCEPTED 13 March 2023
PUBLISHED 22 March 2023

CITATION
Chen B, Wei N, Qu T, Zhang L, Li Y, Long X and Lin Y (2023), Research on weighting method of geological hazard susceptibility evaluation index based on apriori Algorithm.
Front. Earth Sci. 11:1127889.
doi: 10.3389/feart.2023.1127889

Research on weighting method of geological hazard susceptibility evaluation index based on apriori Algorithm

Bin Chen[1,2], Na Wei[1], Tianqiang Qu[3]*, Lianzhi Zhang[3], Yingyi Li[1], Xiao Long[3] and Youxi Lin[4]

[1]School of Civil Engineering, Xiangtan University, Xiangtan, Hunan, China, [2]Hunan Provincial Key Laboratory of Geomechanics and Engineering Safety, Xiangtan University, Xiangtan, Hunan, China, [3]Hunan Provincial Territorial Space Survey and Monitoring Institute, Changsha, Hunan, China, [4]School of Civil Engineering, Changsha University, Changsha, Hunan, China

Given the inconsistency between the information value and the weight value in the weighted information value model, a weight model based on the Apriori algorithm is established in this paper to analyze the correlation between the second-level intervals of disaster factors and the susceptibility of geological disasters. The objective weight of the second-level intervals of each index factor is calculated through the mining of association rules by the Apriori algorithm. The subjective uncertainty of the existing second-level factor weighting method is eliminated. Taking the geological disaster data of Xiangtan urban area as an example, 10 evaluation indexes were selected to establish the entropy weight method-information value (EWM-IV) model and the entropy weight method-Apriori algorithm-information value (EWM-Apriori-IV) model to evaluate the geological disaster susceptibility, and the disaster area ratio and the receiver operating characteristic curve (ROC) verification method were used to test and analyze the evaluation results. The results showed that compared with the EWM-IV model, the EWM-Apriori-IV model is used to evaluate the disaster area ratio of high-prone area increased by 58.3%, and the disaster area ratio of low-prone area decreased by 43.1%, the area under the curve (AUC) increased by 7.4%, and the evaluation accuracy was relatively improved compared with the former. This paper proves the rationality and practicability of the weighting method of the geological hazard susceptibility evaluation index based on the Apriori algorithm.

KEYWORDS

apriori algorithm, susceptibility evaluation, weighted information value model, rule of association, second-level factor weight

1 Introduction

Geological hazard susceptibility evaluation is an important link and basis for disaster prevention and reduction (Chen et al., 2005; Ma et al., 2021). Currently, the commonly used susceptibility evaluation models include analytic hierarchy process model (Chung and Fabbri, 1999; Wang et al., 2009; Xu et al., 2009), weighted information value model (Wang et al., 2014; Jiao et al., 2019; Alsabhan et al., 2022), logistic regression model (Budimir et al., 2015; Tang and Ma, 2015), artificial neural network model (Nourani et al., 2014), support vector machine model (Kavzoglu et al., 2014), *etc.*, among which the weighted information value model is widely used in the research field of geological disaster

susceptibility evaluation due to its clear physical significance and simple algorithm. Shen H et al. (Shen et al., 2021) established a weighted information value model based on the weight value and information quantity value of each index obtained by the analytic hierarchy process (AHP) and information value model to conduct a comprehensive assessment of landslide susceptibility. Liang L et al. (Liang et al., 2019) used the certainty factor model to determine the first-level weight of each index factor, and then multiplied the information value to establish a weighted information value model to evaluate the susceptibility of geological disasters. Yang P et al. (Yang et al., 2020) multiplied the weight of each first-level index factor determined by the random forest model and the information value of each second-level index factor determined by the information value model respectively, and established the weighted information value model for the evaluation of landslide susceptibility. All of the above studies are based on the establishment of weighted information value model for susceptibility evaluation by obtaining the weight of first-level factors and combining with the second-level interval information value of each factor. However, the problem of inconsistency between the weight value of first-level factors and the information value of second-level factors may occur, thus affecting the accuracy of geological disaster susceptibility evaluation. Aiming at this problem, some scholars (Xie, 2011; Wang et al., 2012) determined the second-level interval weights of indicators based on the trapezoidal fuzzy number with subjective experience for susceptibility evaluation, but the evaluation results were worse than the mathematical statistics model with objective weighting. In view of this, this paper introduces the weighting method of geological disaster susceptibility evaluation index based on Apriori algorithm, and obtains the second-level factor weight value by analyzing objective data. Apriori algorithm was proposed by R.Aglawal et al. based on previous research results of AIS algorithm (Agrawal and Srikant, 1994; Yu, 2004). This algorithm is applicable to transaction database association rule mining and can reflect the interdependence and correlation between one thing and other things through association rules. Wu T et al. (Wu and Niu, 2011) used the Apriori algorithm to dig for the correlation between disaster scale and various single and multiple factors. Li J et al. (Li and Niu, 2013) obtained the relationship between land use type and landslide stability through the Apriori algorithm. Jie Q et al. (Jie et al., 2015) used the Apriori algorithm to excavate the deformation laws of several landslides. The above scholars use Apriori algorithm to dig out the internal connection between geological disasters and first-level disaster factors, and prove the feasibility of Apriori algorithm in the field of geological disasters. Therefore, according to the principle that the Apriori algorithm can mine the commonness of historical disaster data, it is integrated into the weighted information value model and weighted to the second-level interval information value of each evaluation factor to solve the problem that the factor weight value is inconsistent with the caliber of the information value.

Based on this, this paper takes Xiangtan urban area in Hunan Province as the research area, introduces the Apriori algorithm to mine and analyze the association rules between historical disaster data and geological disaster susceptibility, establishes the evaluation index system of geological disaster susceptibility in Xiangtan urban area, determines the objective weights of the second-level intervals of each index factor, and combines with the objective weight of first-level index factor of entropy weight method. The EWM-IV model and the EWM-Apriori-IV model were established respectively to evaluate the susceptibility of geological disasters. The feasibility and accuracy of the weight model based on the Apriori algorithm were demonstrated through accuracy verification and comparative analysis.

2 Materials and methods

2.1 Overview of the study area and data source

Xiangtan urban area, Hunan Province, with a total area of 657.4 km^2, is located in the middle part of the Hengshan Mountain range and belongs to the alluvial plain and red soil terrace on both sides of the Xiangjiang River. The terrain is high and convex in the east and west, and low and concave in the middle. The overall terrain is relatively flat, with an elevation between −82 m and 289 m. The outlying strata in Xiangtan urban area are relatively complete and are characterized by the wide distribution of red beds and pre-Devonian shallow metamorphic clastic rocks and a complex and diverse sedimentary environment. From old to new, Lengjiaxi Group, Banxi Group, Sinian System, Cambrian system, Ordovician system, Devonian system, Carboniferous system, Permian system, Triassic system, Jurassic system, Cretaceous system, Paleogene system and Quaternary system (including alluvium and residual slope layer) are exposed successively. The regional structure can be divided into three types: the Heling-Nanzhushan fault folds belt in the northwest, the Xiangtan fault depression basin in the middle, and the Baimalong-Shuangmazhen fault folds belt in the east. The river system in the territory is mainly composed of the Xiangjiang River and its main tributaries Lianshui River and JinJiang River. Affected by the subtropical humid monsoon climate, it enjoys abundant sunshine, and high temperature and rain in summer. The annual rainfall is between 1,250 and 1,500 mm, mainly in spring and summer, with 68% of the annual rainfall. There are 121 geological disaster spots in Xiangtan urban area (Figure 1), mainly landslide, collapse, and ground collapse disasters. The threatened population is about 2,500, and the potential economic loss is nearly 10,000 yuan. The main data sources of this paper are shown in Table 1.

2.2 Methods

2.2.1 Establishment of weight model based on apriori algorithm

The Apriori algorithm is the most classic algorithm for mining frequent item sets, which can extract association rules from large data sets (Zhang, 2016; Hidayanto et al., 2017). Algorithm steps (Han and Kamber, 2001; Yu, 2004; Li et al., 2020) are shown in Figure 2 below.

Apriori algorithm adopts iterative method to find frequent item sets. In the mining process of association rules, the frequent item sets of k-1 items will be connected all the time to generate k item sets. Then, the minimum support threshold is set by calculating the

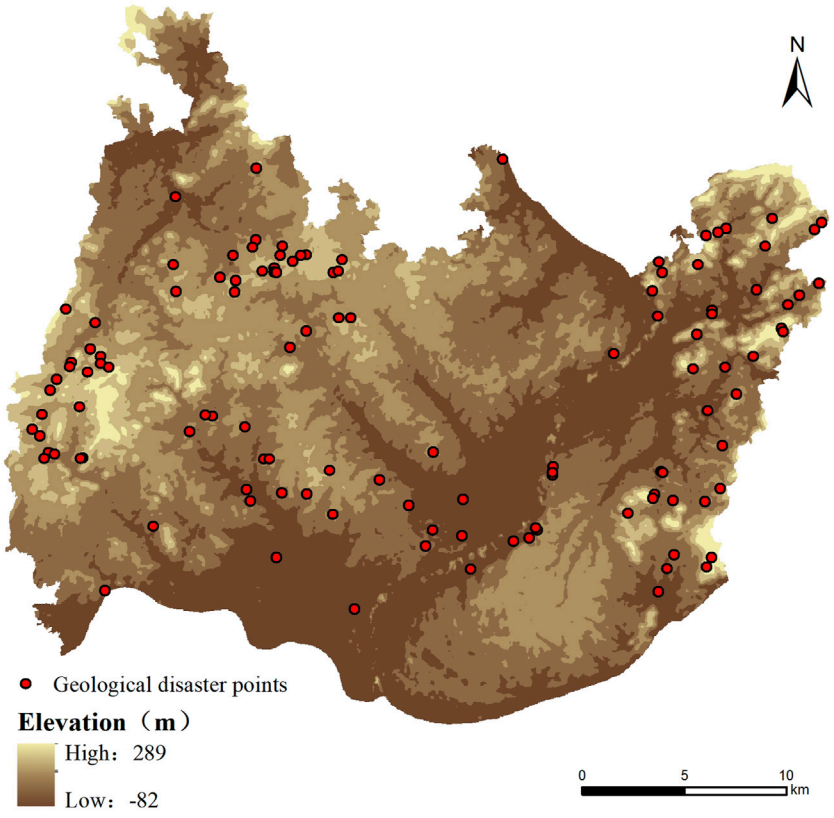


FIGURE 1
Geological hazard distribution map of Xiangtan urban area.

TABLE 1 Basic data of the study area.

Name	Type	Resolution	Extraction data
RS image	Raster	0.5 m	Vegetation coverage
DEM	Raster	1:10000	Altitude, Slope angle, Slope aspect
Engineering geology map	Vector	1:50000	Geologic structure, Lithology
Land use map, Landform type map	Vector	1:50000	Land use pattern, Landform
Fault map, Road map	Vector	1:50000	Distance to faults, Distance to roads
Rainfall data (1990–2019)	Data table	-	Average annual rainfall

support degree of k item sets, namely, the probability of X and Y appearing simultaneously in the data set, and the frequent item sets of k items are obtained by pruning the k item sets that do not meet the threshold. The frequent item sets of k items are searched and iterated layer by layer until there are no new k item sets.

The Apriori algorithm is applied to mine the correlation between the susceptibility of geological disasters and the second-level interval of the disaster factor. The second-level interval of the disaster factor is a data type in the form of Boolean (binary type). According to the requirement of the Apriori algorithm that the data type must be a Boolean value, the objective weight of the second-level interval of each disaster factor reflecting the correlation of geological disaster susceptibility can be obtained. The specific implementation process is as follows:

(1) Storing data sets.

The data set contains historical geological disaster data and all geological data in the study area. Boolean data after data processing, each set in the data set contains a disaster point and the second-level interval to which all corresponding disaster factors belong. Each set represents the occurrence of a disaster point and the emergence of second-level intervals of all disaster factors. The data set is scanned and the Apriori algorithm is run, searching for each transaction until the search result is obtained.

(2) Generate the candidate item set.

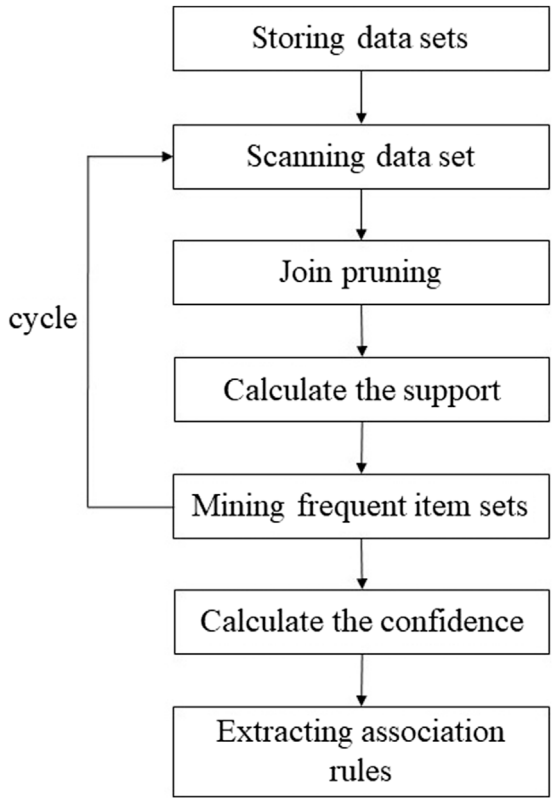


FIGURE 2
Flow chart of the Apriori algorithm.

Scan the transaction set and defines all the second-level intervals in the transaction set and the occurrence of geological hazards as the members of the candidate set. Each member of the candidate item set is an independent item set.

(3) Calculate the weight of the index factor second-level interval.

Eq. 1 was used to calculate the support degree of each member in the candidate set. Support degree of association rule X→Y refers to the probability of the combination of second-level intervals in the geographic data set of the study area and the occurrence of geological disasters at the same time. If the occurrence occurs once, the support degree accumulates once. The support result is the weight of the second-level intervals of each indicator factor.

$$\text{Support}(X \rightarrow Y) = P(X \cap Y) \tag{1}$$

(4) Determine the weight of the index factor's second-level interval.

In the process of mining association rules, by setting the minimum support threshold, the association rules that do not meet the minimum support threshold are pruned to improve the efficiency of the algorithm. If the minimum threshold of the above support is met, it can be determined as the second-level weight of the index factor.

2.2.2 Entropy weight method

As an objective weight method with a good evaluation effect, the entropy weight method (EWM) has been widely used in research and practice of geological disaster susceptibility evaluation (Liang et al., 2010; Jiang et al., 2019). The specific calculation method is as follows (Devkota et al., 2013):

$$FR_{ij} = \frac{b_{ij}}{a_{ij}} \tag{2}$$

$$P_{ij} = \frac{FR_{ij}}{\sum_{j=1}^{n} FR_{ij}} \tag{3}$$

Where FR_{ij} is the occurrence frequency of geological disasters; P_{ij} is frequency density; a_{ij} and b_{ij} are the disaster quantity and regional area in the j second-level interval of the i first level factor, respectively; n is the total number of second-level intervals of the i first-level factor.

The entropy value of the i first-level factor can be expressed as H_i:

$$H_i = -e\sum_{j=1}^{n} P_{ij} \ln P_{ij} \tag{4}$$

Where $e = \frac{1}{\ln n}$ is constant; In order to ensure the $0 \le H_i \le 1$, regulation: if $P_{ij} = 0$, the $\ln P_{ij} = 0$.

Finally, the objective weight w_i of the index first-level factor can be obtained:

$$w_i = \frac{1 - H_i}{N - \sum_{i=1}^{N} H_i} \tag{5}$$

Where N is the total number of index factors.

2.2.3 Information value model

The information value (IV) model is a mathematical statistics method commonly used to analyze data. Through statistical analysis of historical data, the information value of each impact factor is calculated to determine the importance of each impact factor (Fan et al., 2012; Chen et al., 2021). The calculation method is as follows:

$$I_{ij} = \ln \frac{N_{ij}/N}{S_{ij}/S} \tag{6}$$

Where I_{ij} is the information value of the j second-level interval of the i first-level factor; S is the layer area of each evaluation index; S_{ij} is the interval area of the j second-level interval of the i first-level factor; N is the total number of disaster points in the study area; N_{ij} is the number of disaster points in S_{ij} interval.

Combined with the weight model, the weighted information value (WIV) model is constructed. The calculation formula is as follows:

$$I = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(I_{ij} \times Q_{ij}\right) \tag{7}$$

Where I is the total information value of each evaluation unit in the overlay layer; n is the total number of index factors; m is the total number of second-level intervals in the first-level factors of each index; Q_{ij} is the index weight value. If the entropy weight method is used for weighting, Q_{ij} is the first-level weight of each index factor calculated by the entropy weight method. If the entropy weight method-Apriori algorithm is used for weighting, Q_{ij} is the second-level comprehensive weight of each index factor, the objective weight w_i of each first-level factor calculated by the entropy weight method is multiplied with the weight r_{ij} of the corresponding second-level

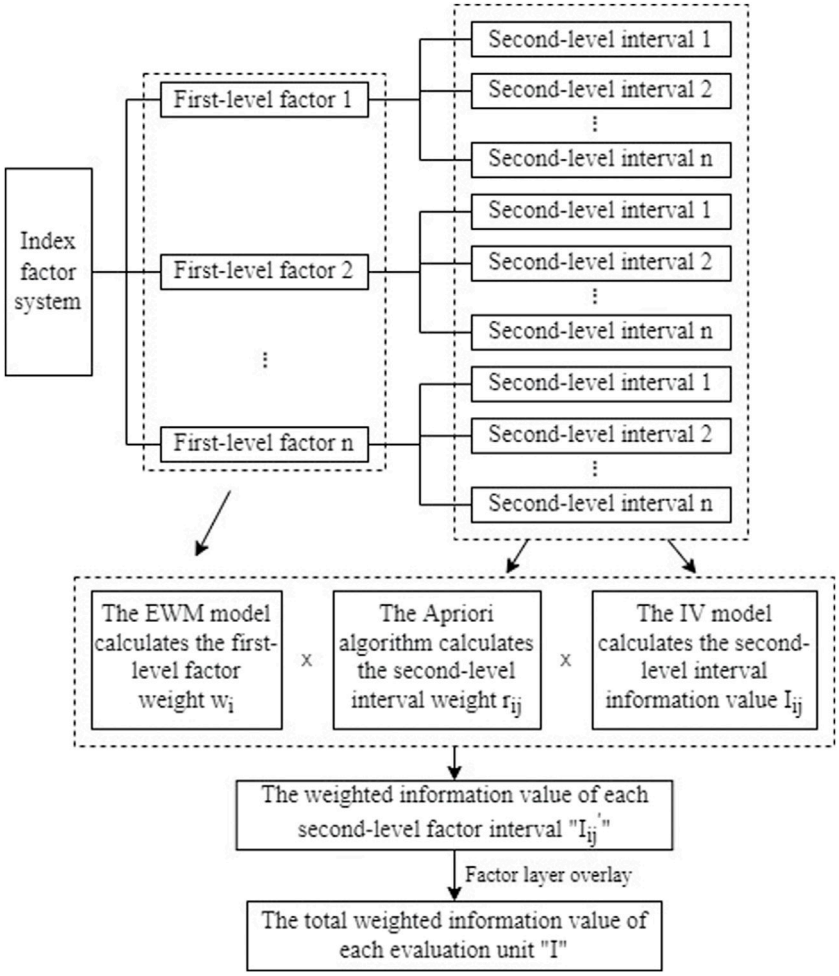


FIGURE 3
Calculation flowchart of EWM-Apriori-IV model.

interval calculated by the Apriori algorithm, as shown in Eq. 8. Figure 3 shows the calculation process of EWM-Apriori-IV model based on the entropy weight method-Apriori algorithm.

$$Q_{ij} = w_i \times r_{ij} \tag{8}$$

3 Results

3.1 Data processing

According to the characteristics of historical geological disaster data in the study area and previous research results (Lan et al., 2002; Shahri et al., 2019), 10 geological disaster susceptibility factors including altitude, slope angle, slope aspect, landform, lithology, vegetation coverage, average annual rainfall, distance to faults, distance to roads and land use pattern were initially selected to evaluate the geological disaster susceptibility in the study area (Meng et al., 2010; Chen et al., 2013; Zhao et al., 2021). With 30 m×30 m grid units as evaluation units, the research area was divided into 907,823 units. In order to facilitate data processing by the Apriori algorithm, each factor needs to be classified to convert into Boolean data. The natural breakpoint method was used to classify each factor in ArcGIS software, and the data characteristics of each factor in the study area were analyzed by adhering to the principle of "similar within the interval, different outside the interval". The partitioning results of each evaluation factor were shown in Figure 4.

According to the historical landslide disaster point and the second-level interval Boolean data of each index factor in the study area (Table 3), the Apriori algorithm is applied for data mining to analyze the association rules that have guiding significance for the construction of geological disaster susceptibility evaluation index system in the study area, as shown in Table 2.

According to Table 2, the confidence level of the association rule "G4, D1, J2, E9, A5, B5, F2″ is the highest, indicating that the index factors contained in the disaster causing condition obtained when mining this association rule should be the favorable candidate factors for the evaluation index system of geological disaster susceptibility in the study area. Meanwhile, four association rules in the table cover all index factors with high

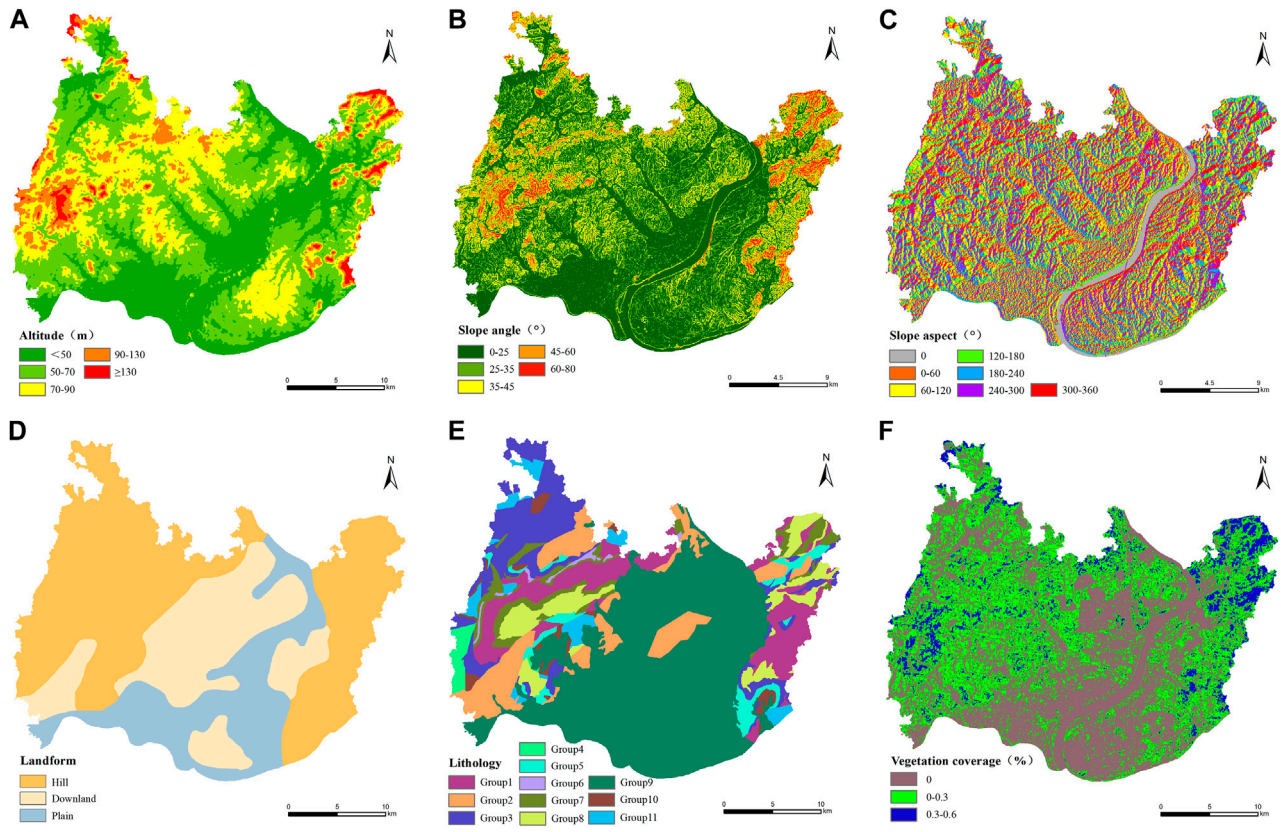


FIGURE 4
(Continued).

confidence, indicating that 10 index factors initially selected, including altitude, slope angle, slope aspect, landform, lithology, vegetation coverage, average annual rainfall, distance to faults, distance to roads, and land use pattern, should be selected as index factors for the evaluation of geological disaster susceptibility in the study area.

3.2 Determination of weight based on the apriori algorithm

Through the analysis and processing of geographical data and historical geological disaster data of the study area, the internal commonness of historical disaster points is explored. According to the above 10 susceptibility evaluation indexes such as altitude, slope angle, slope aspect, lithology, and average annual rainfall, the Apriori algorithm is used to calculate the correlation between the second-level interval of each index factor and the occurrence of geological disasters. To determine the contribution rate (weight) of different second-level intervals of each indicator factor to the geological disasters in the study area (Zhang and Jiang, 2004), that is, to explore the statistical relationship between the point value data of each indicator factor located in different second-level intervals and the distribution of geological disasters in the study area.

The different second-level intervals of each susceptibility evaluation index factor were numbered, that is, the altitude was "1″, and the five second-level factor intervals were "1.1, 1.2, 1.3, 1.4, and 1.5″respectively. The slope angle is classified as "2″, and its 5 second-level factor intervals are respectively "2.1, 2.2, 2.3, 2.4, 2.5″, and so on. Data of all second-level factor intervals of each historical geological disaster point are numbered. According to the implementation process of the Apriori algorithm in Section 2.2.1, the statistical data set of the disaster point is input into the software for program implementation by Python language, and the objective weight value of the second-level interval of each index factor is finally obtained, as shown in Table 3.

3.3 Evaluation of susceptibility

The entropy weight method was used to calculate the first-level objective weight of each evaluation index, and the information value of the second-level factor interval of each evaluation index was obtained by the information value model. The calculation results are shown in Table 3. According to the superposition analysis function of ArcGIS software, combined with the weight results of the above evaluation indicators and the information value calculation results, the EWM-IV model and the EWM-

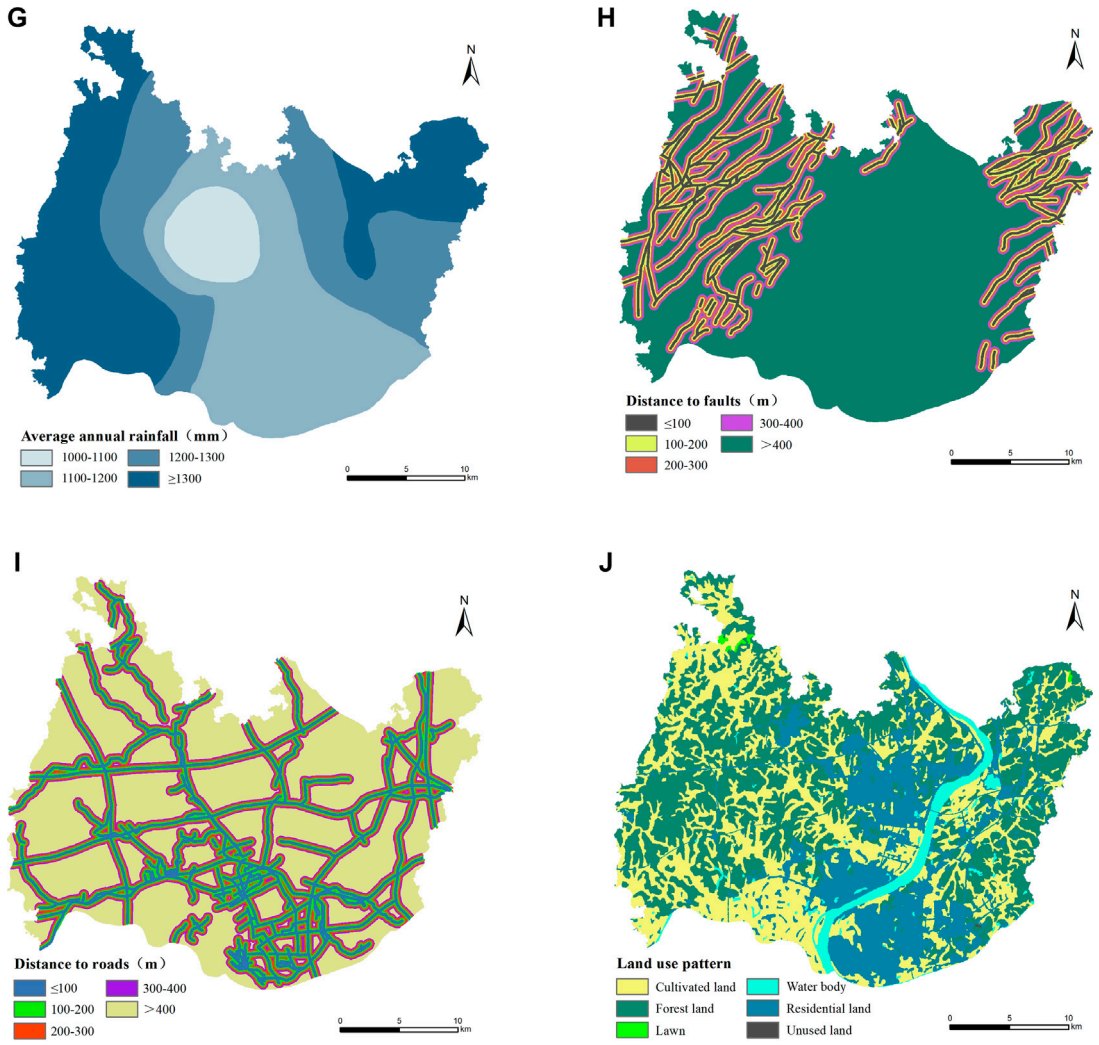


FIGURE 4
(Continued). Zoning map of geological hazard susceptibility assessment factors in Xiangtan urban area **(A)** Altitude; **(B)** Slope angle; **(C)** Slope aspect; **(D)** Landform; **(E)** Lithology; **(F)** Vegetation coverage; **(G)** Average annual rainfall; **(H)** Distance to faults; **(I)** Distance to roads; **(J)** Land use pattern.

TABLE 2 Mining results of association rules.

Serial number	Association rule	confidence coefficient
1	(G4, D1, J2, E9, H5, I5, A5, B5, C3, F2) → Geological disaster occurs	0.73
2	(G4, D1, J2, E9, H5, I5, A5, B2, C3, F2) → Geological disaster occurs	0.72
3	(G4, D1, J2, E9, H5, I5, A5, B5, C3, F1) → Geological disaster occurs	0.72
4	(G4, D1, J2, E9, H5, I5, A1, B5, C3, F2) → Geological disaster occurs	0.71
5	(G4, D1, J2, E9, H5, I5, A5, B5, F2) → Geological disaster occurs	0.82
6	(G4, D1, J2, E9, A5, B5, F2) → Geological disaster occurs	0.87

Apriori-IV model was established according to Eq. 7 respectively to evaluate the susceptibility of geological disasters in the study area. The natural breakpoint method was used to partition the evaluation results of the two models, and the results of geological disaster susceptibility evaluation of Xiangtan urban area were respectively obtained, as shown in Figure 5.

3.4 Comparison of evaluation accuracy

At present, there are disaster area ratio verification methods and ROC curve verification methods to verify the evaluation results of geological disaster susceptibility (Kamp et al., 2008; Bai et al., 2010; Fan et al., 2014). The disaster area ratio

10.3389/feart.2023.1127889

TABLE 3 Calculation results of grading weight and information value (IV) of each evaluation factor.

Index factor	The second-level interval	First level weight	Second level weight	IV
Altitude (m) (A1-A5)	<50	0.2015	0.0075	−0.4487
	50–70		0.0299	−0.3168
	70–90		0.1045	0.3264
	90–130		0.4179	0.7568
	≥130		0.4403	0.6218
Slope angle (°) (B1-B5)	0–25	0.2376	0.0467	0.3826
	25–35		0.2880	−1.5611
	35–45		0.1744	0.5707
	45–60		0.1907	−0.1204
	60–80		0.3002	0.0000
Slope aspect (°) (C1-C6)	<60	0.0116	0.1035	1.5022
	<120		0.1623	−1.0681
	<180		0.2596	−1.4547
	<240		0.1907	−1.9395
	<300		0.1886	−0.7558
	<360		0.0953	−1.8060
Landform (D1-D3)	Hill	0.0410	0.7907	0.4166
	Downland		0.1047	−0.8247
	Plain		0.1047	−0.8513
Lithology (E1-E11)	Slate (1)	0.1685	0.0974	0.6706
	Red clastic rock (2)		0.0000	−0.4590
	Quartzite conglomerate, sandstone (3)		0.1988	0.2673
	Siliceous rock, siliceous shale (4)		0.0933	0.5961
	Carbonate rocks interbedded with clastic rocks (5)		0.0020	0.3173
	Siliceous rock, siliceous slate (6)		0.0933	2.5452
	Shallow metamorphic sandstone sandwiched with slate (7)		0.0020	0.6272
	Sandstone is interbedded with slate (8)		0.0365	0.6946
	Soil body (9)		0.4097	−0.8104
	Limestone, dolomitic limestone with marl (10)		0.0609	−0.7542
	Clastic rocks (11)		0.0061	−1.2896
Vegetation coverage (%) (F1-F3)	0	0.0177	0.3400	−0.5042
	0–0.3		0.5100	0.1912
	0.3–0.6		0.1500	0.7068
Average annual rainfall (mm) (G1-G4)	1,000–1,100	0.1011	0.0116	−0.6563
	1,100–1,200		0.2907	−0.3479
	1,200–1,300		0.2907	0.1833
	≥1,300		0.4070	0.1803
Distance to faults (m) (H1-H5)	≤100	0.0182	0.1860	0.7201

(Continued on following page)

TABLE 3 (*Continued*) Calculation results of grading weight and information value (IV) of each evaluation factor.

Index factor	The second-level interval	First level weight	Second level weight	IV
	100–200		0.1395	0.2456
	200–300		0.0930	0.3223
	300–400		0.1163	0.8306
	>400		0.4651	−0.4579
Distance to roads (m) (I1-I5)	≤100	0.0070	0.1163	−0.0939
	100–200		0.1744	−0.1259
	200–300		0.1395	0.1276
	300–400		0.1047	−0.5607
	>400		0.4651	0.0894
Land use pattern (J1-J6)	Cultivated Land	0.1958	0.0714	0.0081
	Forest land		0.7500	0.2004
	Lawn		0.1303	0.0000
	Waterbody		0.0000	−0.8964
	Residential land		0.0483	−0.3481
	Unused land		0.0000	0.0000

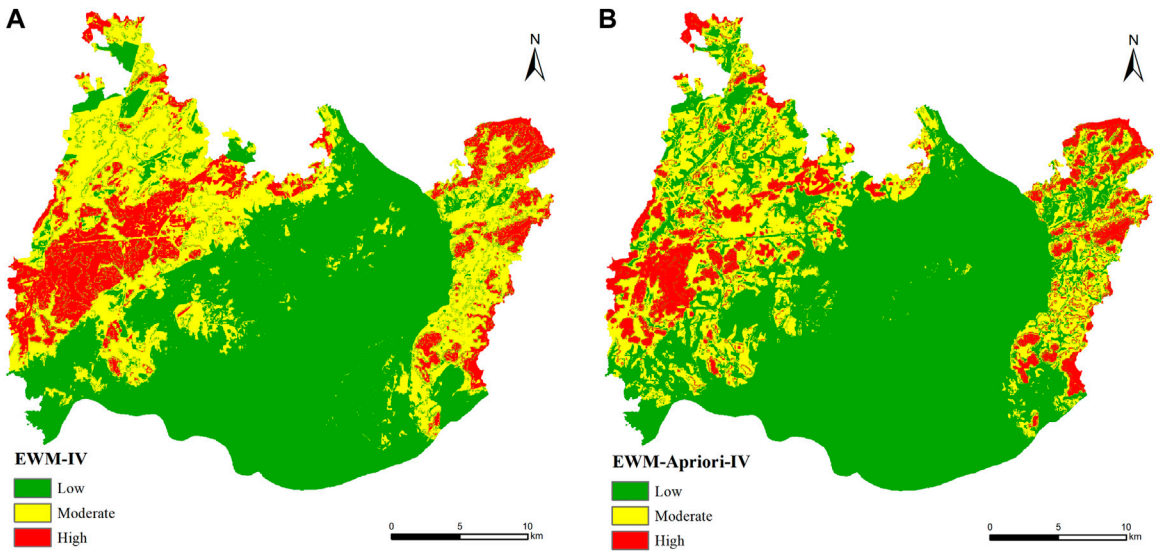


FIGURE 5
Zoning map of geological hazard susceptibility assessment in Xiangtan urban area **(A)** EWM-IV model; **(B)** EWM-Apriori-IV model.

TABLE 4 The disaster area ratio of EWM-IV model evaluation results.

Susceptibility zoning	Proportion of area (%)	Number of disasters	The proportion of disasters (%)	The disaster area ratio
High	13.5	64	52.9	0.721
Moderate	28.8	35	28.9	0.185
Low	57.7	22	18.2	0.058

TABLE 5 The disaster area ratio of EWM-Apriori-IV model evaluation results.

Susceptibility zoning	Proportion of area (%)	Number of disasters	The proportion of disasters (%)	The disaster area ratio
High	10.4	78	64.4	1.141
Moderate	24.1	29	24.0	0.183
Low	65.5	14	11.6	0.033

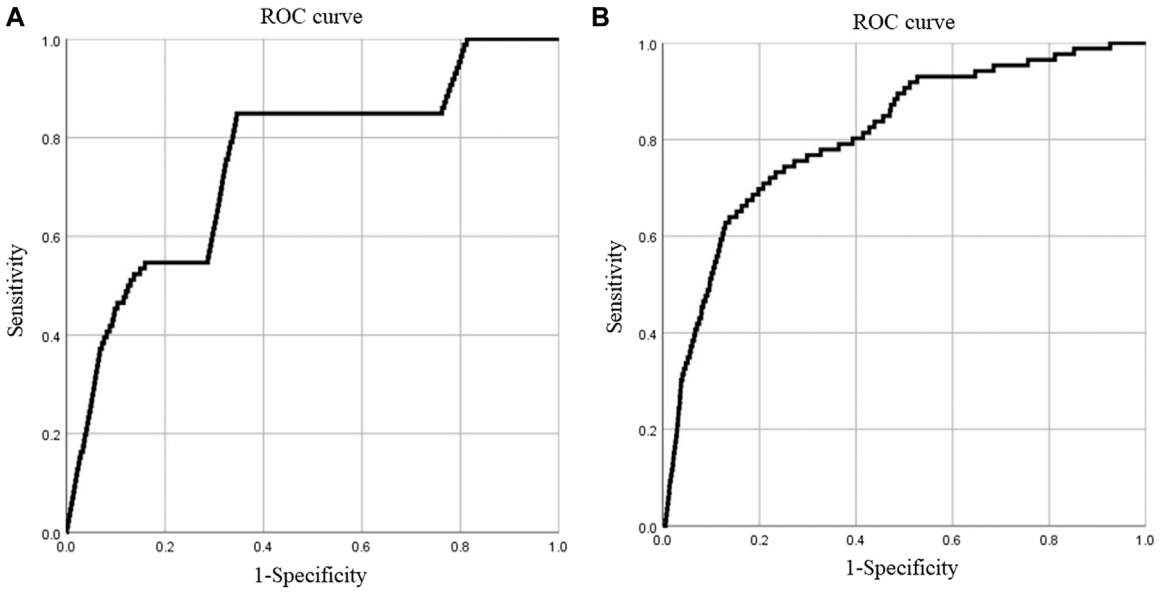


FIGURE 6
ROC curves of different evaluation model. **(A)** EWM-IV model; **(B)** EWM-Apriori-IV model.

verification method compares and verifies the ratio between the number of historical disaster points and the area in each prone area. The larger the disaster area ratio in the high-prone area and the smaller the disaster area ratio in the low-prone area, the more accurate and effective the evaluation results will be (Chen et al., 2019). ROC curve refers to the receiver operating characteristic curve, and the area under the curve (AUC) is used to judge the accuracy of model evaluation results. The higher the AUC value, the better the prediction ability of the model (Pradhan, 2013; Wang et al., 2013).Tables 4, 5 indicates the disaster area ratio statistics of the EWM-IV model and the EWM-Apriori-IV model.

By comparing the statistical results of the disaster area ratio evaluated by the two models, the disaster area ratio of the high-prone areas evaluated by the EWM-Apriori-IV model is 1.141, which is higher than 0.721 of the EWM-IV model, and a relative increase of 58.3%. The disaster area ratio of the low-prone areas evaluated by the EWM-Apriori-IV model is 0.033, lower than the EWM-IV model's 0.058, which is a relative decrease of 43.1%. The results show that the use of the EWM-Apriori-IV model to evaluate the susceptibility of geological disasters in the study area greatly improves the disaster area ratio of the high-prone area and reduces the disaster area ratio of the low-prone area, and the evaluation effect is more accurate and effective.

The ROC curve test was conducted according to the evaluation results of geological disaster susceptibility generated by the two evaluation models. The sensitivity was taken as the vertical coordinate and the specificity was taken as the horizontal coordinate, and the evaluation results were imported into SPSS software for analysis. The results were presented in Figure 6.

The area under the ROC curve (AUC) generated by the two models is both 0.7%-0.9, indicating that the success rate and prediction degree of the results obtained by the two evaluation models used in this paper are between 70% and 90% when evaluating the susceptibility of geological disasters in the study area, with high accuracy. It can be seen from Figure 6 that the AUC values of the EWM-IV model and the EWM-Apriori-IV model are respectively 0.753 and 0.809, that is, the predictive ability of the model is 75.3% and 80.9%, and the latter is 7.4% higher than the former, showing better predictive ability.

4 Discussion

Accuracy of geological hazard susceptibility evaluation is affected by the weighting method of evaluation index. As a commonly used

susceptibility evaluation model, the weighted information value model is established in most studies by combining the first-level index factor weighting method with the second-level interval information value of each factor, while the second-level interval weight of the index factor is rarely analyzed. Therefore, the inconsistent caliber of information value and weight value will be generated by using this model for evaluation, which will affect the evaluation accuracy. In this paper, the Apriori algorithm is introduced to improve the weighted information value model. By analyzing the correlation between the second-level interval of each disaster causing factor and the occurrence of geological disasters, the objective weight of the second-level interval of each evaluation index is determined, and the susceptibility evaluation of geological disasters is completed by combining the second-level interval information value of each index factor. Thus, the problem that the information value is inconsistent with the weight value in the weighted information value model is solved.

According to the evaluation results, the regionalization effect of EWM-Apriori-IV model is better than that of EWM-IV model. Compared with EWM-IV model, the disaster area ratio of high and low prone areas evaluated by EWM-Apriori-IV model increased by 58.3% and decreased by 43.1%, respectively, and the AUC value of the area under ROC curve increased by 7.4%. Therefore, it is proved that the accuracy and rationality of introducing Apriori algorithm to obtain the second-level interval weight of the index factor and combining with the information value model to predict susceptibility. In addition, this paper only analyzes the feasibility of improving the weighted information value model based on Apriori algorithm. For more susceptibility evaluation methods, it is worth further discussion and discussion to combine them with the weighting method of geological hazard susceptibility evaluation index based on Apriori algorithm to carry out index second-level factor weighting and susceptibility evaluation.

5 Conclusion

Taking the Xiangtan urban area of Hunan Province as the research area, this paper selected 10 evaluation indexes, introduced the Apriori algorithm as the weight model of the second-level interval of the evaluation index, constructed two evaluation models to evaluate the susceptibility of geological disasters in the research area, and carried out precision verification and comparative analysis on the evaluation results. The results are as follows.

(1) The Apriori algorithm is introduced to analyze the correlation between the second-level intervals of disaster-causing factors and the susceptibility to geological disasters. A weighted model based on the Apriori algorithm is established to achieve the objective weighting of the second-level intervals of disaster-causing factors in the susceptibility evaluation of geological disasters and to solve the problem of the inconsistency between the information value and the weight value in the weighted information value model.
(2) The EWM-IV model and the EWM-Apriori-IV model are established respectively to evaluate the susceptibility of geological disasters in the study area. The results show that: When the weight model based on the Apriori algorithm is used to assign a weight, the accuracy of susceptibility evaluation is significantly increased by 7.4%, and the disaster area ratio of the high-prone area is increased by 58.3%, while the disaster area ratio of the low-prone area is reduced by 43.1%, indicating that the EWM-Apriori-IV model is more accurate and rational in evaluation.
(3) According to the susceptibility evaluation results of the EWM-Apriori-IV model, the study was divided into high, medium, and low susceptibility areas. The high-prone area covers 68.4 km^2, accounting for 10.4% of the Xiangtan urban area, and contains 78 geological disaster points, accounting for 64.4% of the total geological disaster points. The middle-prone area is 158.4 km^2, accounting for 24.1% of the Xiangtan urban area, including 29 geological disaster points, accounting for 24.0% of the total geological disaster points. The low-prone area covers 430.6 km^2, accounting for 65.5% of the Xiangtan urban area, and contains 14 geological disaster points, accounting for 11.6% of the total geological disaster points.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material further inquiries can be directed to the corresponding author.

Author contributions

BC: Conceptualization, methodology, paper writing and revision; NW: Data collation and analysis, paper writing; TQ: Resource provision; LZ, YiL, XL, and YoL: All participated in the revision, reading and approval of manuscripts, as well as the supervision and management of the process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1127889/full#supplementary-material

References

Agrawal, R., and Srikant, R. (September, 1994). Fast Algorithms for mining association rules, Proceedings of the 20th international conference on very large data bases. IEEE. San Francisco, CA, USA,

Alsabhan, A. H., Singh, K., Sharma, A., Alam, S., Pandey, D. D., Rahman, S. A. S., et al. (2022). Landslide susceptibility assessment in the Himalayan range based along Kasauli – parwanoo road corridor using weight of evidence, information value, and frequency ratio. *J. King Saud. Univ–Sci.* 34 (2), 101759. doi:10.1016/j.jksus.2021.101759

Bai, S., Wang, J., Lü, G., Zhou, P., Hou, S., and Xu, S. (2010). GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorghlogy* 115 (1), 23–31. doi:10.1016/j.geomorph.2009.09.025

Budimir, M. E. A., Atkinson, P. M., and Lewis, H. G. (2015). A systematic review of landslide probability mapping using logistic regression. *Landslides* 12 (3), 419–436. doi:10.1007/s10346-014-0550-5

Chen, P., Qiao, J., Peng, Z., Xie, K., and Yu, H. (2013). Screening of debris flow risk factors and risk evaluation based on rank correlation. *Rock Soil Mech.* 34 (5), 1409–1415. doi:10.16285/j.rsm.2013.05.015

Chen, G., Tang, M., Zhou, H., Qu, F., Li, Y., and Xin, Y. (2019). Dynamic risk assessment method of geological hazard of linear engineering in mountainous area and its application. *J. Disaster Prev. Mitig. Eng.* 39 (3), 524–532. doi:10.13409/j.cnki.jdpme.2019.03.020

Chen, J., Yang, Z., and Liu, H. (2005). Landslide susceptibility zoning and its probabilistic forecast. *Chin. J. Rock Mech. Eng.* 13, 2392–2396. doi:10.3321/j.issn:1000-6915.2005.13.032

Chen, L., Guo, H., Gong, P., Yang, Y., Zuo, Z., and Gu, M. (2021). Landslide susceptibility assessment using weights-of-evidence model and cluster analysis along the highways in the Hubei section of the Three Gorges Reservoir Area. *Comput. Geosciences* 156, 104899. doi:10.1016/j.cageo.2021.104899

Chung, C., and Fabbri, A. G. (1999). Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Eng. Remote Sens.* 65 (12), 1389–1399.

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., et al. (2013). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* 65 (1), 135–165. doi:10.1007/s11069-012-0347-6

Fan, L., Hu, R., Zeng, F., Wang, S., and Zhang, X. (2012). Application of weighted information value model to landslide susceptibility assessment—a case study of enshi city, hubei province. *J. Eng. Geol.* 20 (4), 508–513. doi:10.3969/j.issn.1004-9665.2012.04.005

Fan, Q., Ju, N., Xiang, X., and Huang, J. (2014). Landslides hazards assessment with weights of evidence—A case study in guizhou, China. *J. Eng. Geol.* 22 (3), 474–481. doi:10.13544/j.cnki.jeg.2014.03.017

Han, J., and Kamber, M. (2001). *Data mining concepts and techniques*. Beijing, China: Higher Education Press.

Hidayanto, B. C., Muhammad, R. F., Kusumawardani, R. P., and Syafaat, A. (2017). Network intrusion detection systems analysis using frequent item set mining algorithm FP-max and Apriori. *Procedia Comput. Sci.* 124, 751–758. doi:10.1016/j.procs.2017.12.214

Jiang, Y., Wu, G., Zhao, Z., and He, L. (2019). Channel selection method for mountain railway based on terr-ain and geomorphological parameters. *J. Eng. Geol.* 27 (4), 903–913. doi:10.13544/j.cnki.jeg.xq2019010

Jiao, Y., Zhao, D., Ding, Y., Liu, Y., Xu, Q., Qiu, Y., et al. (2019). Performance evaluation for four GIS-based models purposed to predict and map landslide susceptibility: A case study at a world heritage site in southwest China. *CATENA* 183, 104221. doi:10.1016/j.catena.2019.104221

Jie, Q., Shi, B., Luo, W., Fei, B., and Sun, Y. (2015). Analysis of DFOS-based association rules on multi-fields data monitoring of slope. *J. Eng. Geol.* 23 (6), 1146–1152. doi:10.13544/j.cnki.jeg.2015.06.015

Kamp, U., Growley, B. J., Khattak, G. A., and Owen, L. A. (2008). GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology* 101 (4), 631–642. doi:10.1016/j.geomorph.2008.03.003

Kavzoglu, T., Sahin, E. K., and Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11 (3), 425–439. doi:10.1007/s10346-013-0391-7

Lan, H., Wu, F., Zhou, C., and Wang, S. (2002). Sensitivity analysis of landslide factors in xiaojiang basin of yunnan province based on GIS. *Chin. J. Rock Mech. Eng.* 21 (10), 1500–1506. doi:10.3321/j.issn:1000-6915.2002.10.014

Li, B., Li, Y., and Cai, M. (2020). Research on key factors of water environment for cyanobacteria growth in Taihu Lake based on data mining. *J. Hohai Univ. Nat. Sci.* 48 (6), 506–513. doi:10.3876/j.issn.1000-1980.2020.06.004

Li, J., and Niu, R. (2013). Association rule mining and using between stability and land use of landslides—Taking area from zigui to badong in three gorges reservoir area as an example. *Comput. Eng. Appl.* 49 (6), 234–239. doi:10.3778/j.issn.1002-8331.1107-0608

Liang, G., Xu, W., and Tan, X. (2010). Application of extension theory based on entropy weight to rock quality evaluation. *Rock Soil Mech.* 31 (2), 535–540. doi:10.16285/j.rsm.2010.02.024

Liang, L., Liu, Y., Tang, Z., Zou, Q., and Li, J. (2019). Geologic hazards susceptibility assessment based on weighted information value—a case study in luding county, sichuan province. *Bull. Soil Water Conservation* 39 (6), 176–182+321. doi:10.13961/j.cnki.stbctb.2019.06.026

Ma, Z., and Mei, G. (2021). Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth Sci. Rev.* 223, 103858. doi:10.1016/j.earscirev.2021.103858

Meng, F., Li, G., Li, M., Ma, J., and Wang, Q. (2010). Application of stepwise discriminant analysis to screening evaluation factors of debris flow. *Rock Soil Mech.* 31 (9), 2925–2929. doi:10.16285/j.rsm.2010.09.003

Nourani, V., Pradhan, B., Ghaffari, H., and Sharifi, S. S. (2014). Landslide susceptibility mapping at Zonouz Plain,Iran using genetic programming and comparison with frequency ratio,logistic regression,and artificial neural network models. *Nat. Hazards* 71 (1), 523–547. doi:10.1007/s11069-013-0932-3

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. &Geosciences* 51, 350–365. doi:10.1016/j.cageo.2012.08.023

Shahri, A. A., Spross, J., Johansson, F., and Larsson, S. (2019). Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *CATENA* 183, 104225. doi:10.1016/j.catena.2019.104225

Shen, H., Dong, Y., Yang, M., Shi, C., and Yang, X. (2021). Assessment on landslide susceptibility in Gansu Province based on AHP and information quantity method. *Res. Soil Water Conservation* 28 (6), 412–419. doi:10.13869/j.cnki.rswc.2021.06.034

Tang, C., and Ma, G. (2015). Study on the zoning method of small regional geological disaster susceptibility based on geomorphic unit. *Sci. Geogr. Sin.* 35 (1), 91–98. doi:10.13249/j.cnki.sgs.2015.01.011

Wang, J., Guo, J., Wang, W., and Fang, L. (2012). Application and comparison of weighted linear combination model and logistic regression model in landslide susceptibility mapping. *J. Central South Univ. Sci. Technol.* 43 (5), 1932–1939.

Wang, J., Yin, K., and Xiao, L. (2014). Landslide susceptibility assessment based on GIS and weighted information value: A case study of wanzhou district, three gorges reservoir. *Chin. J. Rock Mech. Eng.* 33 (4), 797–808. doi:10.13722/j.cnki.jrme.2014.04.012

Wang, L. J., Sawada, K., and Moriguchi, S. (2013). Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Comput. Sci.* 57, 81–92. doi:10.1016/j.cageo.2013.04.006

Wang, W. D., Xie, C. M., and Du, X. G. (2009). Landslides susceptibility mapping in Guizhou province based on fuzzy theory. *Min. Sci. Technol.* 19 (3), 399–404. doi:10.1016/S1674-5264(09)60075-2

Wu, T., and Niu, R. (2011). Association rule mining of geological hazard factors in Badong area of Three Gorges Reservoir Area. *Comput. Eng. Appl.* 47 (11), 240–244. doi:10.3778/j.issn.1002-8331.2011.11.068

Xie, C. (2011). Landslides hazard susceptibility evaluation based on weighting model. *J. Central South Univ. Sci. Technol.* 42 (6), 1772–1779.

Xu, C., Dai, F., Yao, X., Chen, J., Tu, X., Sun, Y., et al. (2009). GIS-based landslide susceptibility assessment using analytical hierarchy process in Wenchuan earthquake region. *Chin. J. Rock Mech. Eng.* 28 (2), 3978–3985. doi:10.3321/j.issn:1000-6915.2009.z2.100

Yang, P., Wang, N., Guo, Y., and Ma, X. (2020). Assessment of landslide susceptibility in Lintong District using weighted information value model. *J. Yangtze River Sci. Res. Inst.* 37 (9), 50–56. doi:10.11988/ckyyb.20190726

Yu, W. (2004). Implementation of data mining based on Apriori algorithm using VB. *Comput. Eng.* 2, 196–197. doi:10.3969/j.issn.1000-3428.2004.02.075

Zhang, L., and Jiang, S. (2004). A data-driven weighting model for regional landslide sensitivity evaluation and its application. *Hydrogeology Eng. Geol.* 6, 33–36. doi:10.16030/j.cnki.issn.1000-3665.2004.06.007

Zhang, L. (2016). *Python data analysis and mining practice*. Beijing, China: China Machine Press. .

Zhao, Z., Liu, Z., and Xu, C. (2021). Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random forest, CF-svm and CF-rf models. *Front. Earth Sci.* 9. 589630, doi:10.3389/feart.2021.589630

frontiers | Frontiers in Earth Science

TYPE Review
PUBLISHED 24 March 2023
DOI 10.3389/feart.2023.1147427

OPEN ACCESS

EDITED BY
Wei Zhao,
Institute of Mountain Hazards and Environment (CAS), China

REVIEWED BY
Zheng Han,
Central South University, China
Reza Derakhshani,
Utrecht University, Netherlands

*CORRESPONDENCE
Zhifang Zhao,
zhaozhifang@ynu.edu.cn

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 18 January 2023
ACCEPTED 21 February 2023
PUBLISHED 24 March 2023

CITATION
Zhao X, Zhao Z, Huang F, Huang J, Yang Z, Chen Q, Zhou D, Fang L, Ye X and Chao J (2023), Application of environmental variables in statistically-based landslide susceptibility mapping: A review.
Front. Earth Sci. 11:1147427.
doi: 10.3389/feart.2023.1147427

Application of environmental variables in statistically-based landslide susceptibility mapping: A review

Xin Zhao[1,2,3], Zhifang Zhao[4]*, Faming Huang[5], Jiangcheng Huang[2], Zhiquan Yang[6], Qi Chen[4], Dingyi Zhou[2], Liuyang Fang[1,3], Xian Ye[1,3] and Jiangqin Chao[2]

[1]Broadvision Engineering Consultants Co., Ltd., Kunming, China, [2]Institute of International Rivers and Eco-Security, Yunnan University, Kunming, China, [3]Yunnan Key Laboratory of Digital Communications, Kunming, China, [4]School of Earth Sciences, Yunnan University, Kunming, China, [5]School of Infrastructure Engineering, Nanchang University, Nanchang, China, [6]School of Public Security and Emergency Management, Kunming University of Science and Technology, Kunming, China

Environmental variables are crucial factors affecting the development and distribution of landslides, and they also provide vitally important information for statistically-based landslide susceptibility mapping (SLSM). The acquisition and utilization of appropriate and the most influential environmental variables and their combinations are crucial for improving the quality of SLSM results. However, compared with the construction of SLSM models based on machine learning, the acquisition and utilization of high-quality environmental variables have received very little attention. In order to further clarify the research status of the application of environmental variables and possible development directions in future research, this study systematically analyzed the application of environmental variables in SLSM. To this end, a literature database was constructed by collecting 261 peer-reviewed articles (from 2002 to 2021) on SLSM from the Web of Science and CNKI platform (www.cnki.net) based on the keywords of "landslide susceptibility" and "environmental variable." We found that existing methods for determining environmental variables do not consider the regional representativeness and geomorphological significance of the variables. We also found that at present, environmental variables are utilized generally without the realization and understanding of their spatial heterogeneity. Accordingly, this study raises two major scientific issues: 1) Effective identification of important environmental variables required in SLSM. 2) Effective representation of the spatial heterogeneity of environmental variables in SLSM modeling. From the perspective of the identification of dominant variables and their geospatial pattern of heterogeneity, targeted solutions for future research are also preliminarily discussed, including the method for identifying dominant variables from qualitative and quantitative perspectives and SLSM model construction considering the specific geospatial patterns. In addition, the applicability and limitation of the mentioned methods are discussed.

KEYWORDS

literature review, landslide, susceptibility zonation, statistically-based method, environmental variable, geomorphological significance, spatial heterogeneity

1 Introduction

Landslide susceptibility refers to the likelihood of a landslide occurring in a certain area under the influence of local environmental conditions (Brabb, 1984). Applying the ideology that the "past and present are the keys to the future" (Furlani and Ninfo, 2015), statistically-based landslide susceptibility mapping (SLSM) is one of the most widely used methods in landslide susceptibility assessment. This method focuses on building the functional relationship between landslide environmental variables and existing landslide inventories, and further realize the quantitative or qualitative assessment of the spatial possibility of landslide occurrence under a set of geological/geographical environmental conditions (Guzzetti et al., 2005; 2006a; 2006b; Reichenbach et al., 2018). SLSM can be used to effectively predict and identify locations prone to landslides on the regional scale (Alcántara-Ayala et al., 2022), and its results can provide indispensable information for various tasks, such as landslide risk assessment, formulation of disaster reduction measures, and land use planning (van Westen et al., 2008; Erener et al., 2016; Huang et al., 2019; Zhang et al., 2022). In recent years, it has gradually become a hot topic of geological hazard research on a global scale (Chen et al., 2020; Huang et al., 2021a).

Since the mid-1970s, research on SLSM has made significant progress. According to 250 slopes and topographic and geological conditions in the southwest of Germany, Neuland (1976) built a landslide prediction model using the statistical method of binary discriminant analysis, which is considered to be the earliest statistical study of landslide susceptibility assessment. Henceforth, scholars worldwide have published hundreds of papers considering different environmental conditions and landslide distribution in different regions, using various statistical methods. These statistical methods are also called data-driven landslide susceptibility models (Lima et al., 2022). By sorting and summarizing relevant literature, these methods can be generally divided into three categories (Table 1): mathematically-based, geography-based and, pattern recognition-based. The model construction under these three categories is based on geostatistical analysis, geospatial analysis, and machine learning, respectively, with each presenting its own advantages and disadvantages.

Researchers are constantly attempting to determine a functional relationship that is more representative of the relationship (assessment model) between environmental variables (independent variables) and landslide inventories (dependent variables) (Carrara, 1983; Guzzetti et al., 1999; Guzzetti et al., 2006b; Budimir et al., 2015; Zhang et al., 2020) from different perspectives. Particularly in recent years, artificial intelligence represented by machine learning (ML) has introduced new tools of function expression and data analysis for SLSM and their combination significantly promotes the development of model construction for landslide susceptibility analysis (Huang et al., 2017; Zhao et al., 2019; Sun et al., 2021; Lima et al., 2022). Gradually, the focus of SLSM research has shifted to the construction and analysis of susceptibility models. Research interest in this field appears to be increasingly shifting towards the testing of new ML models and the comparison of model performance, striving to develop a better and more suitable prediction model. However, through the analysis of the quality of SLSM by applying the susceptibility quality level (SQL) index (Guzzetti et al., 2006a; Guzzetti et al., 2006b), Reichenbach et al. (2018) reported that the number of high-quality research remains small.

Environmental variables are among the most important input information for SLSM models (van Westen et al., 2008). They represent environmental factors that affect the development of landslides and have important and complex effects on the spatial distribution of landslides (Yan et al., 2019). We sorted out the number of environmental variables used in the literature database, which was constructed using 261 peer-reviewed articles (from 2002 to 2021) collected from the Web of Science and CNKI platform (www.cnki.net) according to the following two keywords: "landslide susceptibility" and "environmental variable." The number of such articles showed a significant growth trend (Figure 1) in the past 20 years. In other words, researchers have gradually realized that environmental variables are crucial for the construction of SLSM models (Ercanoglu and Gokceoglu, 2002; Jebur et al., 2014; Kavzoglu et al., 2015; Trigila et al., 2015; Pham et al., 2016; Nguyen et al., 2019). The purpose is to select as many environment variables as possible and include the ones that may be the most influential.

However, the premise is that researchers require some prior knowledge of the main environmental factors affecting landslides in the study area (Crozier, 1986; Zhao et al., 2019), and the selected environmental variables should be closely related to the environmental characteristics of the study area (Pourghasemi et al., 2012; Lima et al., 2022). The analysis of the literature database reveals that this aspect has not received sufficient attention. Researchers appear to be more interested in new modeling techniques, but they pay little attention to selection and application of high-quality environmental variables required in landslide susceptibility modeling. Therefore, in this study, a systematic review based on the literature database was conducted to further clarify the current status of the application of environmental variables in SLSM. On this basis, possible future development ideas of improving the SLSM prediction quality from the perspective of environmental variables are preliminarily explored.

TABLE 1 Statistical methods used for SLSM.

Category	Main method	Literature	Advantage and disadvantage
Mathematically-based	Geostatistical analysis, such as WOE, I, etc.	Wu et al. (2017), Zhang et al. (2022)	Advantage: Simple and effective Disadvantage: Model with subjectivity
Geography-based	Geospatial analysis, such as GIS, GWR, etc.	Clerici et al. (2002), Kamp et al. (2008), Li et al. (2020)	Advantage: Model with geospatial characteristics Disadvantage: Distance is the only measure of the spatial weight
Pattern recognition-based	Machine learning, such as SVM, RF, ANN, etc.	Taner San (2014), Hong et al. (2019), Azarafza et al. (2021), Nikoobakht et al. (2022)	Advantage: Model with strong classification and recognition ability Disadvantage: High model adaptation power leading to overfitting and accuracy limitation

Notes: WOE, weight of evidence; I, information; GIS, geographic information system; GWR, geographically weighted regression; SVM, support vector machine; RF, random forest; ANN, artificial neural network.

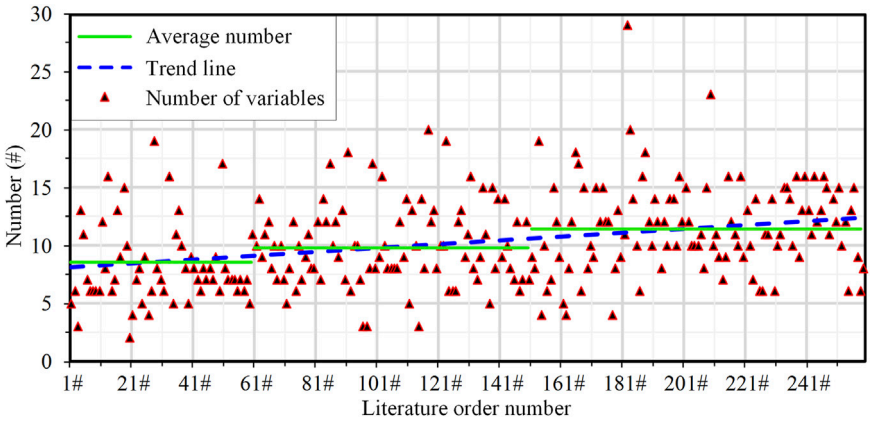


FIGURE 1
Number of environmental variables applied from the literature database. The order numbers are listed in chronological order of the literature.

TABLE 2 Selection methods of environmental variables.

Method	Function	Literature
Empirical discriminant	Experience/cognition → Prior guidance	Pradhan and Lee (2010), Pham et al. (2018), Zhang et al. (2018), Moayedi et al. (2019)
Documentation	Literature documentary records→Direct information reference	Clerici et al. (2002), Ermini et al. (2005), Yalcin (2008), Guo et al. (2015), Hussin et al. (2016)
Data availability	Data availability → Limitation of the selection scope → A clear guidance for selection	Zêzere (2002), Pradhan and Lee (2010), Costanzo et al. (2012), Erener et al. (2016), Juliev et al. (2019)
Comprehensive consideration	Comprehensive consideration →More basis	Myronidis et al. (2016), Hong et al. (2018), Pourghasemi and Rahmati (2018), Hong et al. (2019), Jaafari et al. (2019), Nsengiyumva et al. (2019), Achour and Pourghasemi (2020), Zhu et al. (2021)

2 Acquisition status of environmental variables

The acquisition of environmental variables is a process of searching for inducing and contributing factors that have a certain correlation with the occurrence and distribution of landslides from different aspects such as the natural environment and human activities (Pradhan, 2013; Duan et al., 2022). The ultimate purpose is to ensure that the environmental variables used for SLSM have a logical and effective impact on the prediction of landslide susceptibility in the study area (Ayalew and Yamagishi, 2005; Guzzetti et al., 2012; Jaafari et al., 2019; Pourghasemi et al., 2020). In general, the approach of researchers selecting environmental variables can be summarized into two methods: qualitative selection and quantitative screening.

2.1 Qualitative selection

For a specific study area, the selection of environmental variables is necessary for every SLSM researcher. Through the analysis of the literature database, selection methods of environmental variables could be generally divided into four categories (Table 2): methods based on existing experience and cognition, direct reference from literature records, available existing data of the study area, and comprehensive consideration. At the same time, researchers also consider the accessibility and operability of data in the process of selection from the perspective of qualitative analysis (Lee, 2005). The four methods provide direct and practical guidance and are widely used.

Environmental factors affecting the occurrence and distribution of landslides in a region are complex and diverse (Soeters and van Westen, 1996; Hong et al., 2019), and different regions correspond to different variables (Guzzetti et al., 2006a). According to the analysis of the current status of the qualitative selection of environmental variables in the literature database, although qualitative selection provides a large number of environmental variables, several shortcomings still remain, such as the limited cognition ability of researchers regarding preliminary information on landslides in specific study areas, influence of commonly used environmental variables on the current reference and citation, and difficulties in data acquisition affecting variable selection. Consequently, environmental variables selected for susceptibility assessment may not all possess good expressive ability for the environmental characteristics of the specific study area. For example, for many types of landslides, especially loess or red bed landslides, groundwater variables have a critical impact on the development, movement, and distribution of landslides (Pei et al.,

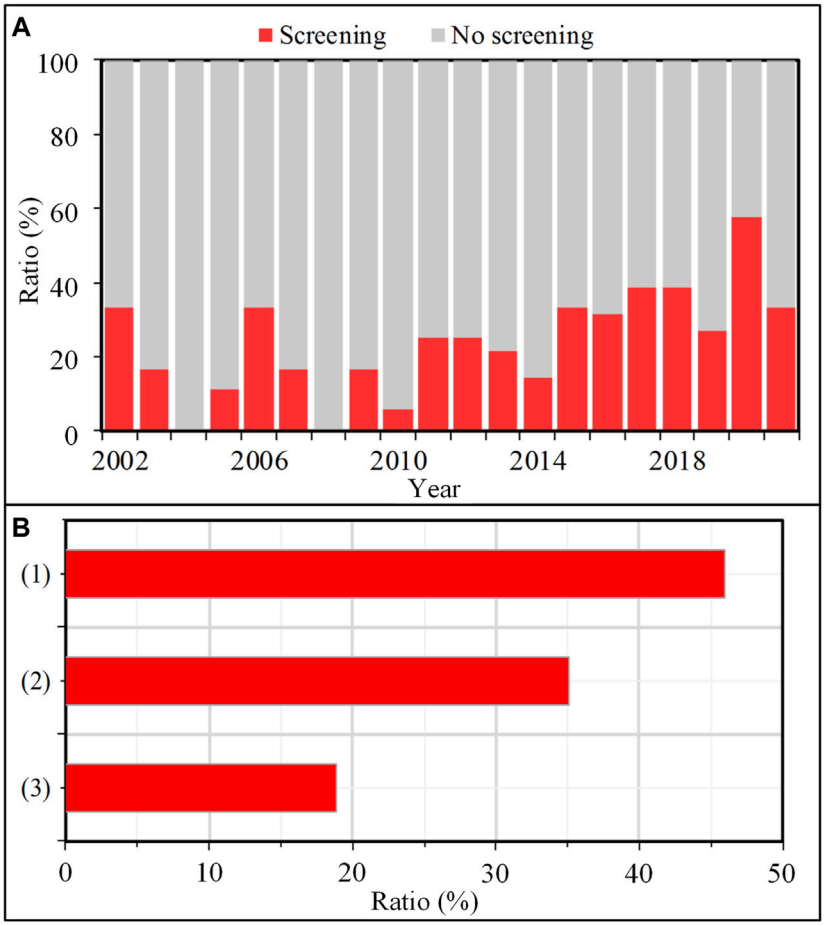


FIGURE 2
(A) Indicates the proportion of screening and non-screening. **(B)** Indicates the proportion of different screening methods.

2018; Peng, 2018; Zhao et al., 2022), but the actual selection and application of groundwater variables in SLSM are rare. In some regions, especially dry and hot valleys, the arid climate conditions have a significant control over the spatial distribution of geological disasters such as landslides and debris flows (Chen et al., 2014; Sun, 2020; Liu et al., 2023). However, drought variables have not received sufficient attention in the current SLSM. Therefore, further in-depth consideration of landslide geological properties and environmental characteristics is required for specific regions in the acquisition of variables.

2.2 Quantitative screening

Some researchers (Lee et al., 2012; Wu et al., 2013; Samia et al., 2017; Lucchese et al., 2021) immediately constructed their susceptibility assessment model after completing the qualitative selection of environmental variables. In contrast, some researchers (He et al., 2019; Al-Najjar and Pradhan, 2021; Chen and Chen, 2021) carried out quantitative screening after completing preliminary selection. Through the analysis of the current status of variable screening in the literature database (Figure 2A), approximately 28% of the studies were found to have further carried out follow-up screening after qualitative selection. Since 2015, the screening of environmental variables has shown a distinct growth trend, and the number of studies with screening reached 39%. In other words, with the continuous deepening of SLSM research, an increasing number of researchers are acknowledging that the screening of environmental variables before constructing the models is conducive to ensuring better conditional independence and higher prediction ability of environmental variables involved in the assessment (Kavzoglu et al., 2015; Chen et al., 2018a; Amato et al., 2019; Hong et al., 2020; Saha et al., 2021).

At present, the screening of environmental variables is primarily based on 1) correlation analysis between variables and landslide distributions, 2) analysis of the conditional independence of variables, and 3) combination analysis (Figure 2B). The first approach accounts for the largest proportion (45.9%) of all screening approaches. In this approach, environmental variables with high correlation are screened by analyzing the relationship between the environmental variables and the landslide spatial distribution. The main screening methods include machine learning (Pradhan and Lee, 2010; Kavzoglu et al., 2015; Amato et al., 2019), correlation analysis (Jebur et al., 2014; Ciurleo et al.,

2016; Steger et al., 2016; Huang et al., 2017; Chen and Li, 2020), and information gain (Chen et al., 2018b; Ngo et al., 2021). The second approach accounts for 35.1%. In this approach, the quality of environmental variables involved in the evaluation are improved by eliminating collinearity and redundancy factors, such that variable screening can be realized. The main methods include the collinearity test (Dou et al., 2019; Zhao, 2020; Azarafza et al., 2021; Chen and Chen, 2021), and chi square test (Hong et al., 2017; Pham et al., 2019). The third approach accounts for 18.9% and essentially involves the combination of different screening methods in the first two approaches (Dou et al., 2020; Hong et al., 2020; Fang et al., 2021).

As different screening methods are widely used, there is no universal and fixed criterion for selecting environmental variables for SLSM (Pradhan and Lee, 2010; Hong et al., 2019). Nevertheless, the original intention of researchers to carry out the screening of environmental variables is consistent, namely, to select variables with high quality as far as possible.

In reality, the development and distribution of landslides are closely related to the geomorphological evolution (Gómez and Kavzoglu, 2005). However, current screening methods do not sufficiently consider regional geomorphic characteristics. Moreover, the geomorphological significance of environmental variables is poorly discussed and justified (Reichenbach et al., 2018). Consequently, the environmental variables used for SLSM may not necessarily be the most suitable and provide effective information required for modeling. Furthermore, some selected variables may even be controversial. Therefore, the acquisition of higher quality environmental variables for SLSM remains to be further explored.

3 Application status of environmental variables

3.1 Quantity of variables

In the study of SLSM, the number of environmental variables used has always been controversial. Some researchers believe that effective assessment requires only a few important variables (Ohlmacher and Davis, 2003; Dahl et al., 2010; Akgün, 2012; Pereira et al., 2012). In contrast, some researchers believe that a large number of variables is more conducive to providing more comprehensive environmental information, implying more accurate landslide susceptibility assessment (Sabatakakis et al., 2013; Dou et al., 2015; Rowden and Aly, 2018; Jaafari et al., 2019). At the same time, many researchers (Remondo et al., 2003; Pradhan and Lee, 2010; Floris et al., 2011; Manzo et al., 2013; Jebur et al., 2014) have proved that the increase of the number of environmental variables cannot guarantee the quality and accuracy of susceptibility assessment, and that it may even be counterproductive. Moreover, in the presence of excessive environmental variables, the susceptibility prediction model may even have multicollinearity, high complexity, and limited explanatory power (Amato et al., 2019).

In this study, the numbers of variables used in different literature were found to be variable, with even large differences (Figure 1). In the literature database, 29 environmental variables were used at most, and 2 at least. The average number is about 10. A series of factors, such as the geological/geographical environment of the study area, disaster development/inducing conditions of the landslides, availability of regional data, and researchers' cognition of landslides in the specific region, will affect the quantity of the environmental variables used for the susceptibility assessment (Guzzetti et al., 1999; Lee and Pradhan, 2007; Nefeslioglu et al., 2011; Felicísimo et al., 2013; Yang G. et al., 2019). As observable, there is no optimal solution for the quantity of environmental variables applied in SLSM. In the actual landslide susceptibility assessment, a corresponding number of representative environmental variables should be selected based on the specific environmental characteristics of the study area.

3.2 Type of variables

In the literature database, 91 types of environmental variables were recorded. According to the attributes, they can be divided into six types: morphological, geological, hydrological, land cover, climatic, and human activity (Figure 3A). Morphological variables are the most used among all types, accounting for more than 40%, followed by geological variables (18.66%), hydrological (16.2%), land cover (14.53%), and human activity (6.18%). Climatic variables are the least applied, accounting for only 3.84%. According to the similarity criteria of names and meanings of variables (Reichenbach et al., 2018), the six main types of environmental variables can be further subdivided into 23 subcategories. Figure 3B reveals the distribution of each subcategory of environmental variables. On the whole, 7 subcategories account for nearly 55% of the total, namely, slope, formation lithology, aspect, elevation, curvature, distance to fault, and distance to river. Among them, slope is the most widely used category, accounting for more than 10%.

3.2.1 Morphological

The most widely used subtypes of morphological variables are slope, aspect, curvature (plane curvature and profile curvature), and elevation, which offer simple and direct measurement of the terrain. The wide use of these simple terrain variables is attributable to two factors: 1) They can be easily obtained through digital terrain data (usually DEM) (Xu and Xiao, 2013); 2) They are the most direct and effective measures of regional terrain/geomorphology and more effective for susceptibility prediction than other types of environmental variables (Fabbri et al., 2003).

As shown in Figure 3B, the slope variable is the most widely used among all environmental variables (carrara and Guzzetti, 1995; Fabbri et al., 2003; Budimir et al., 2015; Lima et al., 2022). Reichenbach et al. (2018) further pointed out that the slope variable is the most effective for identifying susceptible and non-susceptible areas of landslides. The main reasons are as follows: 1) The terrain slope plays an important role in controlling the balance between the sliding force and the anti-sliding force of the slope (Wu and Sidle, 1995). Compared with gentle slopes, steep slopes require stronger anti sliding force to maintain stability. 2) Different slope conditions have distinct differences in the ability to collect and store surface water and groundwater. To some extent, the slope determines the impact of surface runoff and groundwater level fluctuation on slope stability (Sun, 2020). Although simple

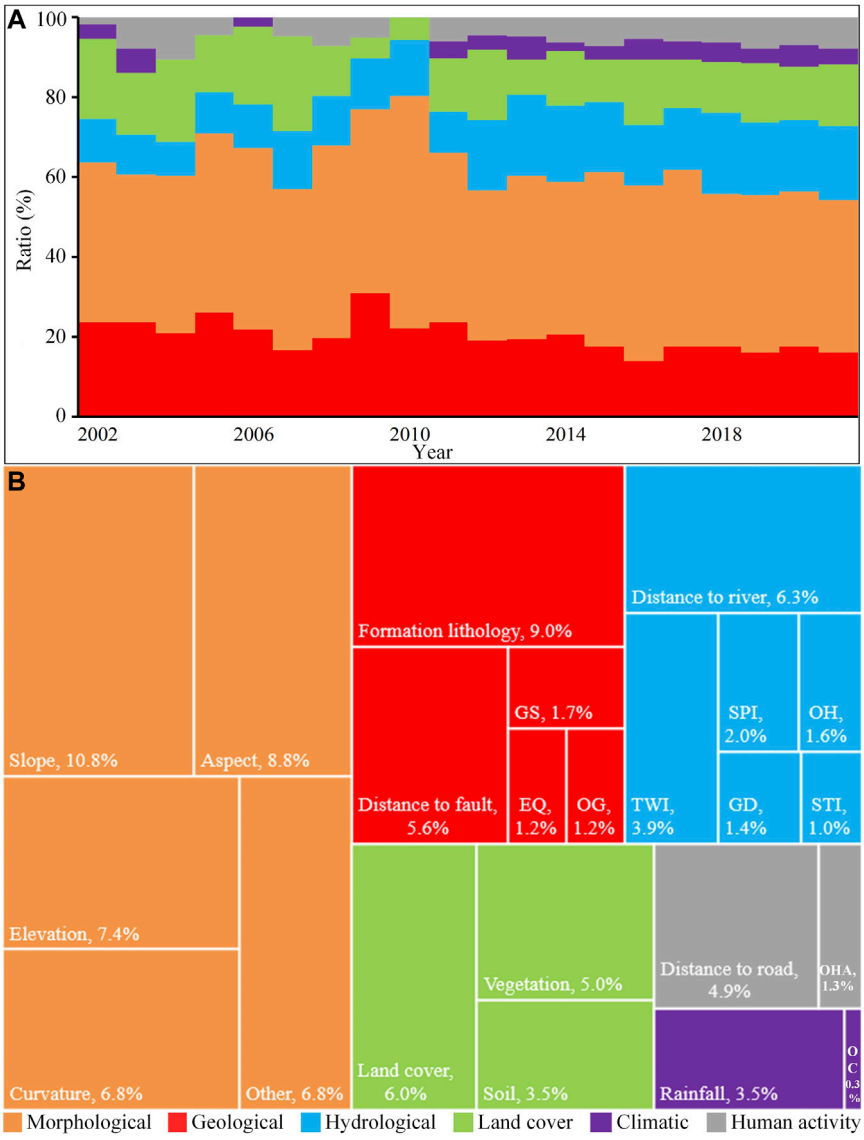


FIGURE 3
Utilization of environmental variables. **(A)** Proportion of different environmental variables in different years. **(B)** Application of different types (main categories and subcategories) of environmental variables. GS, Geological structure; EQ, Earthquake; OG, Other geological; TWI, Topographic wetness index; SPI, Stream power index; GD, Gully density; STI, Sediment transport index; OH, Other Hydrological; OHA, Other human activity; OC, Other climatic.

terrain variables such as aspect, elevation and curvature are also widely used, due to the variable influence of different local conditions in different regions, the rationality of their use has not been effectively demonstrated in theory and practice (Reichenbach et al., 2018), and the rationality of their application in different environmental regions requires further analysis and consideration.

In addition, a type of composite terrain variables also exists, such as surface roughness (SR), topographic position index (TPI), relief degree of land surface, and surface cutting depth. Compared with direct terrain variables, these composite terrain variables can more comprehensively and effectively capture and reflect the overall terrain characteristics of a large area (Lee and Min, 2001; Peng et al., 2014), which can provide strong support for landslide susceptibility prediction. However, in the actual research process, these composite variables have not been widely applied. In the literature database, the SR variable was applied in only 22 articles and the TPI variable in only 13 articles. The main reasons are as follows: first, it is more difficult to obtain composite terrain variables than simple terrain variables; second, the composite variables may have very good applicability in some specific areas (Carrara and Guzzetti, 1995), but their effectiveness in other different environmental heterogeneous areas has not been generally confirmed.

3.2.2 Geological

A series of geological variables (Figure 3B) have been applied in landslide susceptibility modeling, such as formation lithology, distance to fault, and geological structure. Among them, the variables of formation lithology are the most widely used. On the

TABLE 3 Remote sensing means for obtaining geological variables.

RS means	Use	Literature
Aerial photogrammetric survey	Geological structure factor survey and Fault inversion	Ayalew and Yamagishi (2005), Rossi et al. (2010)
Satellite image interpretation	Formation lithology and mineral composition inversion, Missing fault identification and Distance to fault extraction	Gómez and Kavzoglu (2005), Lee (2007), San (2014), Kavzoglu et al. (2015), Pham et al. (2016), Pham et al. (2019), Azarafza et al. (2021), Youssef and Pourghasemi (2021)

one hand, formation lithology is the material basis for landslide development and movement. Different lithologies have different levels of hardness and looseness, shear strength, and weathering resistance, which have a direct impact on slope stability; therefore, formation lithology is considered an important factor affecting landslide susceptibility (Li et al., 2018; Zhu et al., 2021); On the other hand, the variables of formation lithology can be easily obtained from existing regional geological maps (Ciurleo et al., 2017; Yang J. et al., 2019; Fang et al., 2021), which greatly promotes their application in susceptibility modeling. The application frequency of distance to fault in geological variables is only second to that of formation lithology. Faults are the most active components of crustal movement, with nearby rock masses having more developed joint fissures and poor weathering resistance; therefore, geological disasters often occur near large faults (Hong et al., 2015).

As convenient and effective sources of data for obtaining regional geological variables, geological maps have been widely used by researchers (Feng et al., 2016; Zhu et al., 2018; Du et al., 2020; Huang et al., 2021b). Through the literature analysis, the remote sensing characteristics of geological variables were found to be becoming increasingly more prominent with the rapid development of remote sensing technology, Accordingly, remote sensing has gradually become another important tool for the inversion and extraction of geological variables. Many scholars obtained geological variables such as fault, lithology and structure by interpreting aerial photos and satellite images (Table 3), which are also rich sources of geological data for landslide susceptibility assessment.

For some regions with complex geological conditions, the geological environment will show significant differences in different sections, such as different fault distribution densities and regional differences in formation lithology exposure. Therefore, the regional distribution of landslides tends to show significant spatial differences (Lin et al., 2021). However, a notable phenomenon is that although remote sensing means are being used increasingly more frequently to obtain geological variables, only few scholars paid sufficient attention to the spatial heterogeneity characteristics reflected by geological variables in remote sensing images.

3.2.3 Hydrological

The distance to river is the most widely used hydrological variable (Figure 3B), accounting for 6.35% of all environmental variables. To some extent, it indicates the hydrological properties and saturation characteristics of the regional terrain (Lima et al., 2022), reflects the influence of hydrological factors on slope stability, and indirectly indicates the spatial distribution of landslides. However, in actual SLSM studies, the significance of the "distance" variables in morphological and geological aspects remain uncertain and needs to be further discussed (Reichenbach et al., 2018).

It is worth noting that DEM derived composite variables such as the topographic wetness index (TWI), stream power index (SPI), and sediment transport index (STI) are widely used among hydrological variables. Particularly in recent years, the magnitude and proportion of the application of these hydrological composite variables have shown a significant increasing trend (Figure 4). One of the most important reasons is that these composite variables can effectively represent the hydrological distribution characteristics of water and sediment movement in the region (Lian et al., 2008), and also indirectly reflect the correlation between regional hydrology and other factors, such as pore water pressure and permeability, that affect slope stability (Zhang et al., 2021; Yin et al., 2022). With these advantages, they are being increasingly applied to the prediction and determination of the spatial distribution of landslides.

In nature, different environmental regions have corresponding water resources and water environments (Yu, 2000), and the spatial distribution of hydrological elements is extremely uneven across regions, showing significant heterogeneity (Dong, 2015; Zheng et al., 2022). Therefore, the accurate representation of the spatial heterogeneity of hydrological variables is crucial for landslide susceptibility analysis. However, only few scholars have discussed this issue in detail thus far.

3.2.4 Land cover

Land cover variables and slope stability are closely related. At present, scholars generally select land cover variables from two aspects: vegetation and land use (Figure 3B). Vegetation is one of the important and key factors. As vegetation is the natural protective barrier of mountain slopes, it has the effects of slope reinforcement, reducing water and soil loss, and mitigating mountain disasters (Stokes et al., 2009; Qiu, 2012; Dai, 2021). Therefore, good vegetation coverage is often an important predictor of regional surface stability (Carrara et al., 1991).

In addition, under the influence of natural and human factors, land cover variables are highly prone to change in space, which in turn directly or indirectly affect the spatial distribution of landslides, and ultimately lead to significant spatial differences in landslide susceptibility (Reichenbach et al., 2014). With the advantages of wide coverage and high resolution, remote sensing technology provides an important and effective means for determining the spatial change characteristics of land cover variables (Liu et al., 2010; Zhang, 2016; Li et al., 2021). The analysis of the literature database revealed that although remote sensing means account for a high proportion in the acquisition methods of land cover variables (60% and 70% for land use and vegetation variables, respectively), researchers essentially began using remote sensing representation

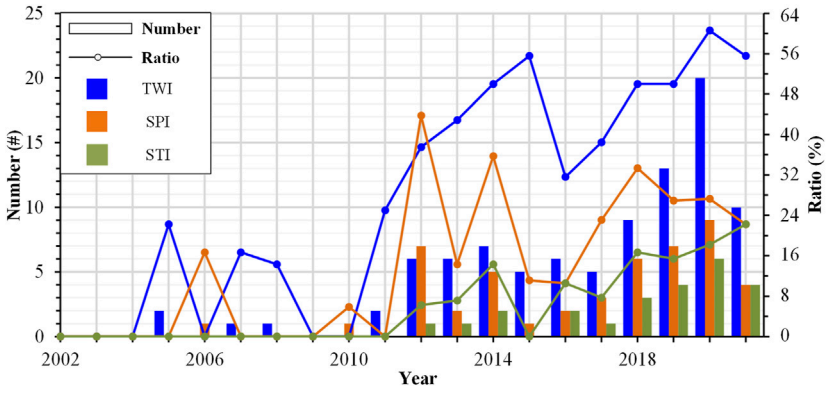


FIGURE 4
Application of composite hydrological variables.

only in recent history (Bordoni et al., 2020; Lin et al., 2021; Saha et al., 2021). The spatial heterogeneity of variables in remote sensing images is rarely included in the prediction and analysis of landslide spatial distribution.

3.2.5 Climatic

Rainfall is one of the most widely used climate variables (Figure 3B). Rainfall can induce slope deformation, which is closely related to the spatial distribution of regional landslides (Li, 2020). Moreover, historical rainfall records are easy to obtain. Therefore, most researchers first consider rainfall when selecting climatic variables (Guo et al., 2015; Chen and Li, 2020; Balogun et al., 2021).

It is generally believed that the spatial distribution of landslides, debris flows, and other geological disasters is primarily controlled by rainfall variables. However, this is not the case in some regions. For example, in the Xiaojiang basin of Yunnan Province, the spatial distribution of debris flow disasters does not simply increase with rainfall, but it has a significant coupling relationship with the drought climate variables (Chen et al., 2014). Rainfall has more of an inductive effect (Yang et al., 2023), while the drought variable, representing regional weathering and material accumulation, has a significant control over the spatial distribution of disasters. For example, in the Sanjiang region at high altitude, eight areas with strong landslide development exhibit large differences in temperature conditions (Yao et al., 2020), but the rainfall variable does not show a significant dominant role. Furthermore, in the Haibalo debris-flow basin in northwest Yunnan, outbreaks of large-scale debris flows are often the result of the joint actions of rainfall and glacier melt water attributable to temperature rise (Zhao et al., 2020). It is observable that the heterogeneity of climatic conditions has an important impact on the development and distribution of geological disasters (Pourkhosravani et al., 2022). Therefore, the significant spatial heterogeneity of climate variables profoundly affects and controls the development and distribution of geological disasters. This phenomenon deserves in-depth analysis in future research because it may provide deeper insight into landslide susceptibility.

3.2.6 Human activity

With the rapid development of the economy and society, human activity variables are becoming increasingly more important in susceptibility assessment. The distance to linear engineering (highways, railways, etc.) is the most commonly used human activity variable. Although some scholars (Bai et al., 2010; Lepore et al., 2012) believe that the rationality of the impact of these variables on the spatial distribution of landslides has not been fully proved, distance variables related to human activities have always been favored by many researchers in the study of SLSM (Pham et al., 2018; Shao et al., 2020). The main reason is that such factors can be directly and conveniently obtained through DEM or topographic data with the help of GIS platforms.

Reichenbach et al. (2018) believe that different forms of linear engineering over slopes (surface cutting or underground crossing, etc.) will have distinct direct impacts on the stabilities of slopes. Moreover, such different forms will also indirectly lead to spatial differences in landslide distribution. With different types of engineering activities, human activity variables exhibit spatial heterogeneity characteristics similar to those of other environmental variables. Therefore, the effects of such heterogeneity characteristics on landslide susceptibility is worth exploring in the future.

4 Discussion and prospect

Compared with the construction of SLSM models based on ML, environmental variables have received very little attention. Nevertheless, their qualities directly determine the scientific nature and accuracy of the results of landslide susceptibility prediction. Therefore, an in-depth exploration on improving the quality of SLSM from the perspective of the acquisition and utilization of the environmental variables will be very important in the future. Through the literature review, this study identified two major scientific issues that need to be solved urgently: 1) Effective identification of important environmental variables required in SLSM. 2) Effective representation of spatial heterogeneity of environmental variables in SLSM modeling.

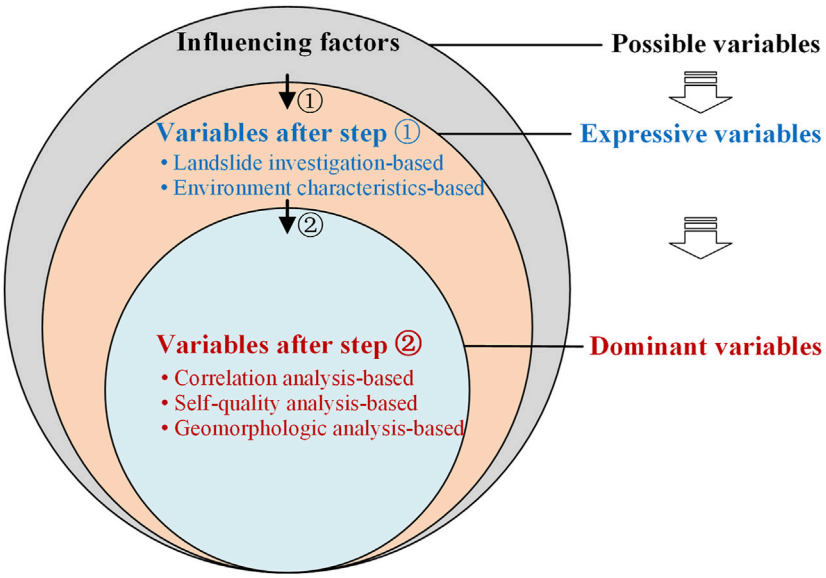


FIGURE 5
Proposed approach for identifying dominant environmental variables. Influencing factors are variables that may affect the spatial distribution of landslides. Step ① indicates the screening procedure of expressive variables based on field investigations and the researchers' recognition of environmental characteristics. Variables after step ① indicate those that are more expressive in environmental characteristics, and are screened from the qualitative perspective. Step ② indicates the identification procedure of dominant variables based on the correlation analysis of landslides and variables, analysis of variable quality, and the influence analysis of variables on geomorphological evolution. Variables after step ② indicate those with strong correlation with landslide, high quality, and significant geomorphological significance screened from a quantitative perspective.

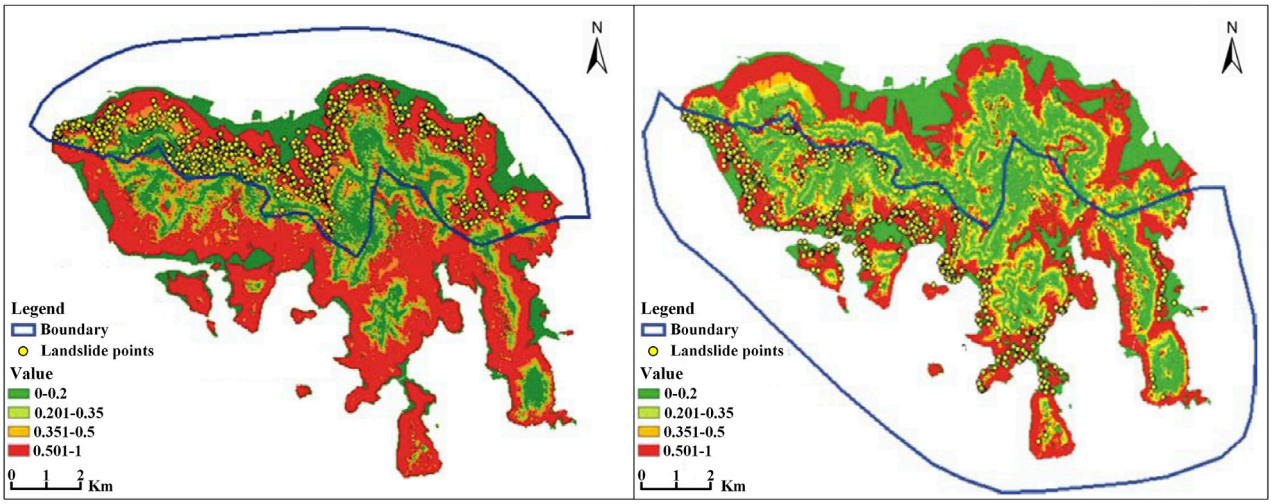


FIGURE 6
Landslides distribution in Hong Kong Island. Landslides are dense in the north and sparse in the south, and the heterogeneity in landslide distributions produces different susceptibility prediction results accordingly (Chau and Chan, 2005).

4.1 Identifying dominant environmental variables

Existing methods of the qualitative selection and quantitative screening of environmental variables have provided direct and practical tools for researchers and also improved the quality of the environmental variables involved in the assessment to a certain extent. However, against the background of the relatively low overall quality of SLSM (Reichenbach et al., 2018) and lack of in-depth consideration of environmental characteristics in variable selection, the method for identifying dominant variables for assessment requires further improvement. First, geo-environmental conditions significantly affect the spatial distribution of landslides (Guzzetti et al., 2005). To ensure sufficient representativeness of the

landslide engineering geological properties and regional environmental characteristics of a specific study area, the selected environmental variables should be able to reflect such important influencing factors. Furthermore, as a common geological disaster in mountain areas, landslides are among the most intense surface processes and external forces of regional geomorphological evolution in the short and long terms (Broeckx et al., 2018), and the spatial distribution of landslides reflects the development and evolution characteristics of the landform to a large extent (Hu, 2019). It can be inferred that the regional geomorphological evolution will also record important environmental information required for predicting the spatial distribution of landslides to a large extent. Therefore, the above two perspectives provide important clues for solving the first scientific issue.

Based on the current research status and existing basis, this study proposes the following method (Figure 5) for identifying the dominant environmental variables for SLSM modeling. Firstly, as many variables with potential effects on the spatial distribution of landslides should be collected as possible based on the methods in Table 1 and the researchers' cognition of the geo-environment of the study area. Subsequently, based on field surveying and the environmental characteristics of the study area, environmental variables with good representativeness of the environmental characteristics should be screened out. Finally, on the basis of traditional methods (Figure 1B), it is necessary to further explore the impact of regional geomorphological evolution on landslide distribution and further identify the geomorphological significance of environmental variables and determine the dominant variables for modeling.

4.2 SLSM analysis considering geospatial patterns

Heterogeneity is a significant spatial distribution characteristic of variables in the natural environment (Cheng et al., 2020; Fu et al., 2021; Liu et al., 2022). As mentioned in Section 3.2, environmental variables required for landslide susceptibility modeling also exhibit significant spatial heterogeneity on the surface, which will lead to different dominant environmental variables and their combinations in different heterogeneous regions. However, considering the current research status, although a few researchers have begun to pay attention to the heterogeneity of environmental variables in landslide susceptibility research (Chang et al., 2022), the realization and understanding of the spatial pattern of the heterogeneity of environmental variables are lacking. Therefore, further exploration is urgently required.

In reality, different regional environmental characteristics will lead to regional differences in the spatial distribution of landslides, which will further lead to different results (Figure 6). According to field surveys, the established spatial distribution of landslides and heterogeneity of environmental variables determine the unique distribution of landslide susceptibility in a given study area, leading to specific spatial patterns. Therefore, in future SLSM research, it is necessary to draw on the idea of research concepts and methods involving the geospatial pattern (Fu, 2014), which would provide a means for solving the second scientific issue. Moreover, SLSM models should be constructed by considering the specific geospatial patterns of the study area. This may be a research direction worth considering for improving the quality of SLSM in the future.

4.3 Applicability and limitation

In fact, as important input information in SLSM studies at the regional scale, the environmental variables involved in the evaluation represent environmental factors affecting the development and distribution of landslides in the study area. Therefore, their acquisition and application should be based on the actual environmental characteristics. Every region has distinctive dominant variables in a specific environment. The identification method discussed in this study (Figure 5) can realize the acquisition of dominant variables from the qualitative (Figure 5 Step ①) and quantitative (Figure 5 Step ②) perspectives. The purpose is to select the most representative and high-quality evaluation factors. However, this method also has some limitations, in that a certain theoretical basis of geomorphology is required and researchers are required to have a deep understanding of environmental characteristics.

The geospatial pattern represents the difference of regional environmental characteristics, which is directly related to the distribution pattern of landslide susceptibility. In fact, the spatial heterogeneity of environmental variables required for SLSM cannot be fully reflected for all regions. For example, the SLSM analysis method mentioned in this study may not be applicable to areas with small area and insignificant environmental differences. In contrast, it may have good reference value for areas with significant differences in environmental characteristics such as landform, geology and climate.

5 Conclusion

Based on the literature database composed of 261 peer-reviewed articles (from 2002 to 2021), this paper systematically summarizes the status of the acquisition and utilization of environmental variables used in SLSM. Moreover, possible research directions in the future are explored.

(1) As one of the important input information of SLSM models, environmental variables significantly affect the accuracy of susceptibility predictions. However, compared with the construction of ML-based susceptibility models, the acquisition and utilization of high-quality environmental variables have received very little attention.
(2) This review revealed that although the current acquisition methods of environmental variables provide direct and practical guidance for researchers to some extent, the environmental variables still lack representativeness of the landslide engineering geological properties and environmental characteristics of specific regions and their geomorphological significance. In other words, the environmental variables used in susceptibility assessment are not all dominant factors. Accordingly, one major scientific issue is raised: Effective identification of important environmental

variables required in SLSM. In view of this, an approach for identifying dominant environmental variables is preliminarily proposed. Nevertheless, the identification method requires a deep understanding of environmental characteristics and a certain theoretical basis of geomorphology.

(3) This review also revealed that the current approach of utilizing various types of environmental variables is generally not based on the acknowledgment and understanding of the spatial heterogeneity of variables. Accordingly, another major scientific issue is raised: Effective representation of the spatial heterogeneity of environmental variables in SLSM modeling. Given that the established distribution of landslides and the heterogeneity of environmental variables constitute a specific spatial pattern, the concepts and methods of the geospatial pattern (Fu, 2014) should be fully taken into consideration for areas with significant environmental heterogeneity in future SLSM research.

Author contributions

XZ: literature database construction and original draft writing. ZZ, FH, JH, and ZY: review and editing. QC, DZ, XY, and JC: investigation and data analysis. LF: paper revision and language polishing.

Funding

This work is supported by the Yunnan Academic and Technical Leader Reserve Talent Project (202305AC160071), the List of Key Science and Technology Projects in the Transportation Industry of the Ministry of Transport in 2021 (Grant No. 2021-MS4-105), the Technological Innovation Plan Project of Yunnan Communications Investment & Construction Group Co., Ltd. (Grant No. YCIC-YF-2022-07), the Science and Technology Project of Yunnan Traffic Planning Design Institute Co., Ltd. (Grant No. ZL-2021-03), the Yunnan Key Laboratory of Digital Communications (Grant No. 202205AG070008), the National Natural Science Foundation of China (Grant No. 41872251), the Plateau Mountain Ecology and Earth's Environment Discipline Construction Project (Grant No. C1762101030017) and the Joint Foundation Project between Yunnan Science and Technology Department and Yunnan University (Grants Nos. C176240210019 and 2019FY003017), the Yunnan Provincial Department of Education Science Research Fund Project (Grant No. 2023Y0196). The authors declare that this study received funding from Yunnan Communications Investment & Construction Group Co., Ltd. and Yunnan Traffic Planning Design Institute Co., Ltd. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

Authors XZ, LF, and XY were employed by the company Broadvision Engineering Consultants Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Achour, Y., and Pourghasemi, H. R. (2020). How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geosci. Front.* 11 (3), 871–883. doi:10.1016/j.gsf.2019.10.001

Akgün, A. (2012). A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision and likelihood ratio methods: Case study at izmir, Turkey. *Landslides* 9, 93–106. doi:10.1007/s10346-011-0283-7

Al-Najjar, H. A. H., and Pradhan, B. (2021). Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. *Geosci. Front.* 12, 625–637. doi:10.1016/j.gsf.2020.09.002

Alcántara-Ayala, I., Cui, P., and Pasuto, A. (2022). Disaster risk reduction in mountain areas: A research overview. *J. Mt. Sci.* 19 (6), 1487–1494. doi:10.1007/s11629-022-7487-2

Amato, G., Eisank, C., Castro-Camilo, D., and Lombardo, L. (2019). Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment. *Eng. Geol.* 260, 105237. doi:10.1016/j.enggeo.2019.105237

Ayalew, L., and Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the kakuda-yahiko mountains, central Japan. *Geomorphology* 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P. M., and Derakhshani, R. (2021). Deep learning-based landslide susceptibility mapping. *Sci. Rep.* 11, 24112. doi:10.1038/s41598-021-03585-1

Bai, S., Wang, J., Lü, G., Zhou, P., Hou, S., and Xu, S. (2010). GIS-Based logistic regression for landslide susceptibility mapping of the zhongxian segment in the three gorges area, China. *Geomorphology* 115 (1), 23–31. doi:10.1016/j.geomorph.2009.09.025

Balogun, A. L., Rezaie, F., Pham, Q. B., Gigović, L., Drobnjak, S., Aina, T. A., et al. (2021). Spatial prediction of landslide susceptibility in western Serbia using hybrid support vector regression (SVR) with GWO, BAT and COA algorithms. *Geosci. Front.* 12, 101104. doi:10.1016/j.gsf.2020.10.009

Bordoni, M., Galanti, Y., Bartelletti, C., Persichillo, M. G., Barsanti, M., Giannecchini, R., et al. (2020). The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models. *Catena* 193, 104630. doi:10.1016/j.catena.2020.104630

Brabb, E. E. (1984). Innovative approaches to landslide hazard and risk mapping. *Proc. 4th Int. Symp. Landslides* 1, 307–324.

Broeckx, J., Vanmaercke, M., Duchateau, R., and Poesen, J. (2018). A data-based landslide susceptibility map of africa. *Earth-Sci. Rev.* 185, 102–121. doi:10.1016/j.earscirev.2018.05.002

Budimir, M. E. A., Atkinson, P. M., and Lewis, H. G. (2015). A systematic review of landslide probability mapping using logistic regression. *Landslides* 12 (3), 419–436. doi:10.1007/s10346-014-0550-5

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., and Reichenbach, P. (1991). GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf. Proc. Land.* 16 (5), 427–445. doi:10.1002/esp.3290160505

Carrara, A., and Guzzetti, F. (1995). Advances in natural technological hazard research series. in *Geographical information systems in assessing natural hazards* (Dordrecht, Netherlands: Kluwer Academic Publisher), 5, 342.

Carrara, A. (1983). Multivariate models for landslide hazard evaluation. *Math. Geol.* 15 (3), 403–426. doi:10.1007/BF01031290

Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S. R., Huang, J., et al. (2022). Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. *J. Rock Mech. Geotech.* doi:10.1016/j.jrmge.2022.07.009

Chau, K. T., and Chan, J. (2005). Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong island. *Landslides* 2 (4), 280–290. doi:10.1007/s10346-005-0024-x

Chen, N., Lu, Y., Zhou, H., Deng, M., and Han, D. (2014). Combined impacts of antecedent earthquakes and droughts on disastrous debris flows. *J. Mt. Sci.* 11 (6), 1507–1520. doi:10.1007/s11629-014-3080-7

Chen, W., and Li, Y. (2020). GIS-Based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena* 195, 104777. doi:10.1016/j.catena.2020.104777

Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., et al. (2018b). Landslide susceptibility modelling using GIS-based machine learning techniques for chongren county, jiangxi province, China. *Sci. Total Environ.* 626, 1121–1135. doi:10.1016/j.scitotenv.2018.01.124

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., et al. (2018a). GIS-Based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena* 164, 135–149. doi:10.1016/j.catena.2018.01.012

Chen, X., and Chen, W. (2021). GIS-Based landslide susceptibility assessment using optimized hybrid machine learning methods. *Catena* 196, 104833. doi:10.1016/j.catena.2020.104833

Chen, Z., Ye, F., Fu, W., Ke, Y., and Hong, H. (2020). The influence of DEM spatial resolution on landslide susceptibility mapping in the baxie river basin, NW China. *Nat. Hazards* 101 (3), 853–877. doi:10.1007/s11069-020-03899-9

Cheng, S., Peng, P., Zhang, H., and Lu, F. (2020). Review of interpolation, reconstruction and prediction methods for heterogeneous and sparsely distributed geospatial data. *Geomatics Inf. Sci. Wuhan Univ.* 45 (12), 1919–1929. doi:10.13203/j.whugis20200488

Ciurleo, M., Calvello, M., and Cascini, L. (2016). Susceptibility zoning of shallow landslides in fine grained soils by statistical methods. *Catena* 139, 250–264. doi:10.1016/j.catena.2015.12.017

Ciurleo, M., Cascini, L., and Calvello, M. (2017). A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. *Eng. Geol.* 223, 71–81. doi:10.1016/j.enggeo.2017.04.023

Clerici, A., Perego, S., Tellini, C., and Vescovi, P. (2002). A procedure for landslide susceptibility zonation by the conditional analysis method. *Geomorphology* 48, 349–364. doi:10.1016/S0169-555X(02)00079-X

Costanzo, D., Rotigliano, E., Irigaray, C., Jiménez-Perálvarez, J. D., and Chacón, J. (2012). Factors selection in landslide susceptibility modelling on large scale following the GIS matrix method: Application to the river beiro basin (Spain). *Nat. Hazards Earth Syst. Sci.* 12, 327–340. doi:10.5194/nhess-12-327-2012

Crozier, M. J. (1986). *Landslides: Causes, consequences and environment*. London: Croom Helm Pub.

Dahl, M. -P. J., Mortensen, L. E., Veihe, A., and Jensen, N. H. (2010). A simple qualitative approach for mapping regional landslide susceptibility in the Faroe Islands. *Nat. Hazards Earth Syst. Sci.* 10, 159–170. doi:10.5194/nhess-10-159-2010

Dai, Z. (2021). *Study on controlling factors of rainfall induced shallow landslide in well vegetated area: A case study of liziyuan forest farm in qinzhou district, gansu Province*. Beijing: Beijing Forestry University. doi:10.26949/d.cnki.gblyu.2021.000723

Dong, Z. (2015). On features of five dominant ecological components of aquatic ecosystem. *Water Resour. Hydropower Eng.* 46 (6), 41–47. doi:10.13928/j.cnki.wrahe.2015.06.008

Dou, J., Yamagishi, H., Pourghasemi, H. R., Yunus, A. P., Song, X., Xu, Y., et al. (2015). An integrated artificial neural network model for the landslide susceptibility assessment of osado island, Japan. *Nat. Hazards* 78, 1749–1776. doi:10.1007/s11069-015-1799-2

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., et al. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the izu-oshima volcanic island, Japan. *Sci. Total Environ.* 662, 332–346. doi:10.1016/j.scitotenv.2019.01.221

Dou, J., Yunus, A. P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., et al. (2020). Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Sci. Total Environ.* 720, 137320. doi:10.1016/j.scitotenv.2020.137320

Du, J., Glade, T., Woldai, T., Chai, B., and Zeng, B. (2020). Landslide susceptibility assessment based on an incomplete landslide inventory in the jilong valley, tibet, Chinese himalayas. *Eng. Geol.* 270, 105572. doi:10.1016/j.enggeo.2020.105572

Duan, Y., Tang, J., Liu, Y., Gao, X., and Duan, Y. (2022). Spatial sensitivity evaluation of loess landslide in liulin county, shanxi based on random forest. *Sci. Geogr. Sin.* 42 (2), 343–351. doi:10.13249/j.cnki.sgs.2022.02.016

Ercanoglu, M., and Gokceoglu, C. (2002). Assessment of landslide susceptibility for a landslide-prone area (north of yenice, NW Turkey) by fuzzy approach. *Environ. Geol.* 41, 720–730. doi:10.1007/s00254-001-0454-2

Erener, A., Mutlu, A., and Düzgün, H. S. (2016). A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Eng. Geol.* 203, 45–55. doi:10.1016/j.enggeo.2015.09.007

Ermini, L., Catani, F., and Casagli, N. (2005). Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66, 327–343. doi:10.1016/j.geomorph.2004.09.025

Fabbri, A. G., Chung, C-J. F., Cendrero, A., and Remondo, J. (2003). Is prediction of future landslides possible with a GIS? *Nat. Hazards* 30 (3), 487–503. doi:10.1023/B:NHAZ.0000007282.62071.75,

Fang, Z., Wang, Y., Peng, L., and Hong, H. (2021). A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int. J. Geogr. Inf. Sci.* 35 (2), 321–347. doi:10.1080/13658816.2020.1808897

Felicísimo, A., Cuartero, A., Remondo, J., and Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslide* 10, 175–189. doi:10.1007/s10346-012-0320-1

Feng, H., Zhou, A., Yu, J., Tang, X., Zheng, J., Chen, X., et al. (2016). A comparative study on plum-rain-triggered landslide susceptibility assessment models in west zhejiang province. *Earth Sci.* 41 (3), 403–415. doi:10.3799/dqkx.2016.032

Floris, M., Iafelice, M., Squarzoni, C., Zorzi, L., De Agostini, A., and Genevois, R. (2011). Using online databases for landslide susceptibility assessment: An example from the veneto region (northeastern Italy). *Nat. Hazards Earth Syst. Sci.* 11, 1915–1925. doi:10.5194/nhess-11-1915-2011

Fu, B. (2014). The integrated studies of geography: Coupling of patterns and processes. *Acta Geogr. Sin.* 69 (8), 1052–1059. doi:10.11821/dlxb201408002

Fu, B., Zhao, W., and Piao, S. (2021). *Physical geography Frontiers*. Beijing: Science Press.

Furlani, S., and Ninfo, A. (2015). Is the present the key to the future? *Earth-Sci. Rev.* 142, 38–46. doi:10.1016/j.earscirev.2014.12.005

Gómez, H., and Kavzoglu, T. (2005). Assessment of shallow landslide susceptibility using artificial neural networks in jabonosa river basin, Venezuela. *Eng. Geol.* 78, 11–27. doi:10.1016/j.enggeo.2004.10.004

Guo, C., Montgomery, D. R., Zhang, Y., Wang, K., and Yang, Z. (2015). Quantitative assessment of landslide susceptibility along the xianshuihe fault zone, Tibetan plateau, China. *Geomorphology* 248, 93–110. doi:10.1016/j.geomorph.2015.07.012

Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central Italy. *Geomorphology* 31, 181–216. doi:10.1016/S0169-555X(99)00078-1

Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., and Cardinali, M. (2006a). Landslide hazard assessment in the collazzone area, umbria, central Italy. *Nat. Hazard. Earth Sys.* 6, 115–131. doi:10.5194/nhess-6-115-2006

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K. -T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* 112, 42–66. doi:10.1016/j.earscirev.2012.02.001

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M. (2006b). Estimating the quality of landslide susceptibility models. *Geomorphology* 81, 166–184. doi:10.1016/j.geomorph.2006.04.007

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72, 272–299. doi:10.1016/j.geomorph.2005.06.002

He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., et al. (2019). Landslide spatial modelling using novel bivariate statistical based naïve bayes, RBF classifier, and RBF network machine learning algorithms. *Sci. Total Environ.* 633, 1–15. doi:10.1016/j.scitotenv.2019.01.329

Hong, H., Ilia, I., Tsangaratos, P., Chen, W., and Xu, C. (2017). A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the wuyuan area, China. *Geomorphology* 290, 1–16. doi:10.1016/j.geomorph.2017.04.002

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., et al. (2018). Landslide susceptibility mapping using J48 decision tree with AdaBoost, bagging and rotation forest ensembles in the guangchang area (China). *Catena* 163, 399–413. doi:10.1016/j.catena.2018.01.005

Hong, H., Liu, J., and Zhu, A-X. (2020). Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. *Sci. Total Environ.* 718, 137231. doi:10.1016/j.scitotenv.2020.137231

Hong, H., Miao, Y., Liu, J., and Zhu, A-X. (2019). Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena* 176, 45–64. doi:10.1016/j.catena.2018.12.035

Hong, H., Pradhan, B., Xu, C., and Bui, D. T. (2015). Spatial prediction of landslide hazard at the yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* 133, 266–281. doi:10.1016/j.catena.2015.05.019

Hu, S. (2019). *Spatial pattern of landslide in loess plateau and its influence on geomorphologic evolution*. Xi'an: Northwest University.

Huang, F., Chen, J., Tang, Z., Fan, X., Huang, J., Zhou, C., et al. (2021a). Uncertainties of landslide susceptibility prediction due to different spatial resolutions and different proportions of training and testing datasets. *Chin. J. Rock Mech. Eng.* 40 (6), 1155–1169. doi:10.13722/j.cnki.jrme.2020.1119

Huang, F., Wang, Y., Dong, Z., Wu, L., Guo, Z., and Zhang, T. (2019). Regional landslide susceptibility mapping based on grey relaional degree model. *Earth Sci.* 44 (2), 664–676. doi:10.3799/dqkx.2018.175

Huang, F., Ye, Z., Jiang, S., Huang, J., Chang, Z., and Chen, J. (2021b). Uncertainty study of landslide susceptibility prediction considering the different attribute interval numbers of environmental factors and different data-based models. *Catena* 202, 105250. doi:10.1016/j.catena.2021.105250

Huang, F., Yin, K., Huang, J., Gui, L., and Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng. Geol.* 229, 11–22. doi:10.1016/j.enggeo.2017.04.013

Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., et al. (2016). Different landslide sampling strategies in a grid-based bivariate statistical susceptibility model. *Geomorphology* 253, 508–523. doi:10.1016/j.geomorph.2015.10.030

Jaafari, A., Panahi, M., Pham, B. T., Shahabi, H., Bui, D. T., Rezaie, F., et al. (2019). Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* 175, 430–445. doi:10.1016/j.catena.2018.12.033

Jebur, M. N., Pradhan, B., and Tehrany, M. S. (2014). Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sens. Environ.* 152, 150–165. doi:10.1016/j.rse.2014.05.013

Juliev, M., Mergili, M., Mondal, I., Nurtaev, B., Pulatov, A., and Hübl, J. (2019). Comparative analysis of statistical methods for landslide susceptibility mapping in the bostanlik district, Uzbekistan. *Sci. Total Environ.* 653, 801–814. doi:10.1016/j.scitotenv.2018.10.431

Kamp, U., Growley, B. J., Khattak, G. A., and Owen, L. A. (2008). GIS-Based landslide susceptibility mapping for the 2005 kashmir Earthquake region. *Geomorphology* 101, 631–642. doi:10.1016/j.geomorph.2008.03.003

Kavzoglu, T., Sahin, E. K., and Colkesen, I. (2015). Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Eng. Geol.* 192, 101–112. doi:10.1016/j.enggeo.2015.04.004

Lee, S. (2007). Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. *Environ. Geol.* 52, 615–623. doi:10.1007/s00254-006-0491-y

Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* 26 (7), 1477–1491. doi:10.1080/01431160412331331012

Lee, S., Hwang, J., and Park, I. (2012). Application of data-driven evidential belief functions to landslide susceptibility mapping in jinbu, korea. *Catena* 100, 15–30. doi:10.1016/j.catena.2012.07.014

Lee, S., and Min, K. (2001). Statistical analysis of landslide susceptibility at yongin, korea. *Env. Geol.* 40 (9), 1095–1113. doi:10.1007/s002540100310

Lee, S., and Pradhan, B. (2007). Landslide hazard mapping at selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4, 33–41. doi:10.1007/s10346-006-0047-y

Lepore, C., Kamal, S. A., Shanahan, P., and Bras, R. L. (2012). Rainfall-induced landslide susceptibility zonation of Puerto Rico. *Environ. Earth Sci.* 66 (6), 1667–1681. doi:10.1007/s12665-011-0976-1

Li, C., Zhuang, D., He, J., and Wen, K. (2021). Spatiotemporal variations in remote sensing phenology of vegetation and its responses to temperature change of boreal forest in tundra-taiga transitional zone in the eastern siberia. *Acta Geogr. Sin.* 76 (7), 1634–1648. doi:10.11821/dlxb202107005

Li, G., Yang, X., Li, H., and Le, Q. (2018). Dataset of geo-hazards in fengjie mapsheet (1:50,000), chongqing. *Geol. China* 45 (S1), 160–169. doi:10.12029/gc2018Z110

Li, H. (2020). *Accounting for spatial autocorrelation into regional landslide susceptibility assessment*. Wuhan: Wuhan University. doi:10.27379/d.cnki.gwhdu.2020.001211

Li, Y., Liu, X., Han, Z., and Dou, J. (2020). Spatial proximity-based geographically weighted regression model for landslide susceptibility assessment: A case study of qingchuan area, China. *Appl. Sci.* 10, 1107. doi:10.3390/app10031107

Lian, G., Guo, X., Fu, B., and Hu, C. (2008). Spatial variability and prediction of soil nutrients on A county scale on the loess plateau: A case study of hengshan county, shaanxi province. *Acta Pedol. Sin.* 45 (4), 577–584.

Lima, P., Steger, S., Glade, T., and Murillo-García, F. G. (2022). Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. *J. Mt. Sci.* 19 (6), 1670–1698. doi:10.1007/s11629-021-7254-9

Lin, Q., Lima, P., Steger, S., Glade, T., Jiang, T., Zhang, J., et al. (2021). National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. *Geosci. Front.* 12, 101248. doi:10.1016/j.gsf.2021.101248

Liu, F., Zhu, A-X., and Pei, T. (2010). Application of high temporal resolution satellite remote sensing in identifying soil texture patterns. *J. Geo-information Sci.* 15 (5), 733–740.

Liu, Q., Liu, H., Zhang, Y., Wu, H., and Deng, M. (2022). A spatio-temporal prediction method of large-scale ground subsidence considering spatial heterogeneity. *Natl. Remote Sens. Bull.* 26 (7), 1315–1325. doi:10.11834/jrs.20211445

Liu, Z., Yang, Z., Chen, M., Xu, H., Yang, Y., Zhang, J., et al. (2023). Research hotspots and Frontiers of Mountain flood disaster: Bibliometric and visual analysis. *Water* 15, 673. doi:10.3390/w15040673

Lucchese, L. V., de Oliveira, G. G., and Pedrollo, O. C. (2021). Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using artificial neural networks. *Catena* 198, 105067. doi:10.1016/j.catena.2020.105067

Wu, H., Zhang, X., Qiao, Y., Liang, Y., Zhang, Y., Yang, S., et al. (2017). Landslide susceptibility evaluation based on certain factor and weight of evidence: a case study in the Longkaikou to Qina section of Jinshajiang watershed. *Geoscience* 31 (6), 1269–1277. Available from: https://www.webofscience.com/wos/alldb/full-record/CSCD:6139846.

Manzo, G., Tofani, V., Segoni, S., Battistini, A., and Catani, F. (2013). GIS techniques for regional-scale landslide susceptibility assessment: The sicily (Italy) case study. *Int. J. Geogr. Inf. Sci.* 27, 1433–1452. doi:10.1080/13658816.2012.693614

Moayedi, H., Mehrabi, M., Mosallanezhad, M., Rashid, A. S. A., and Pradhan, B. (2019). Modification of landslide susceptibility mapping using optimized PSOANN technique. *Eng. Comput.* 35, 967–984. doi:10.1007/s00366-018-0644-0

Myronidis, D., Papageorgiou, C., and Theophanous, S. (2016). Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Nat. Hazards* 81, 245–263. doi:10.1007/s11069-015-2075-1

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., and Gorum, T. (2011). Medium-scale hazard mapping for shallow landslide initiation: The buyukkoy catchment area (cayeli, rize, Turkey). *Landslides* 8, 459–483. doi:10.1007/s10346-011-0267-7

Neuland, H. (1976). A prediction model of landslips. *Catena* 3 (2), 215–230. doi:10.1016/0341-8162(76)90011-4

Ngo, P. T. T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A., et al. (2021). Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci. Front.* 12, 505–519. doi:10.1016/j.gsf.2020.06.013

Nguyen, V. V., Pham, B. T., Vu, B. T., Prakash, I., Jha, S., Shahabi, H., et al. (2019). Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* 10, 157. doi:10.3390/f10020157

Nikoobakht, S., Azarafza, M., Akgün, H., and Derakhshani, R. (2022). Landslide susceptibility assessment by using convolutional neural network. *Appl. Sci.* 12 (12), 5992. doi:10.3390/app12125992

Nsengiyumva, J. B., Luo, G., Amanambu, A. C., Mind'je, R., Habiyaremye, G., Karamage, F., et al. (2019). Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern africa. *Sci. Total Environ.* 659, 1457–1472. doi:10.1016/j.scitotenv.2018.12.248

Ohlmacher, G. C., and Davis, J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng. Geol.* 69, 331–343. doi:10.1016/S0013-7952(03)00069-3

Pei, X., Zhang, S., Huang, R., Zhang, X., Wang, S., Chang, Z., et al. (2018). Study on the centrifuge model test of landslide induced by enrichment of groundwater. *Adv. Eng. Sci.* 50 (5), 55–63. doi:10.15961/j.jsuese.201700983

Peng, D. (2018). *Study on early recognition for potentially loess landslide: A case study at heifangtai terrace, gansu Province, China*. Chengdu: Chengdu University of Technology. doi:10.26986/d.cnki.gcdlc.2018.000010

Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y., and Ye, R. (2014). Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the three gorges area, China. *Geomorphology* 204, 287–301. doi:10.1016/j.geomorph.2013.08.013

Pereira, S., Zêzere, J. L., and Bateira, C. (2012). Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models. *Nat. Hazards Earth Syst. Sci.* 12, 979–988. doi:10.5194/nhess-12-979-2012

Pham, B. T., Pradhan, B., Bui, D. T., Prakash, I., and Dholakia, M. B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of uttarakhand area (India). *Environ. Model. Softw.* 84, 240–250. doi:10.1016/j.envsoft.2016.07.005

Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H., Tran, T., et al. (2019). Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 175, 203–218. doi:10.1016/j.catena.2018.12.018

Pham, B. T., Shirzadi, A., Bui, D. T., Prakash, I., and Dholakia, M. B. (2018). A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the himalayan area, India. *Int. J. Sediment. Res.* 33, 157–170. doi:10.1016/j.ijsrc.2017.09.008

Pourghasemi, H. R., Kornejady, A., Kerle, N., and Shabani, F. (2020). Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence-absence balances on landslide susceptibility mapping. *Catena* 187, 104364. doi:10.1016/j.catena.2019.104364

Pourghasemi, H. R., Mohammady, M., and Pradhan, B. (2012). Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood basin, Iran. *Catena* 97, 71–84. doi:10.1016/j.catena.2012.05.005

Pourghasemi, H. R., and Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* 162, 177–192. doi:10.1016/j.catena.2017.11.022

Pourkhosravani, M., Mehrabi, A., Pirasteh, S., and Derakhshani, R. (2022). Monitoring of maskun landslide and determining its quantitative relationship to

different climatic conditions using D-InSAR and PSI techniques. *Geomat. Nat. Haz. Risk* 13 (1), 1134–1153. doi:10.1080/19475705.2022.2065939

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* 51, 350–365. doi:10.1016/j.cageo.2012.08.023

Pradhan, B., and Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Modell. Softw.* 25, 747–759. doi:10.1016/j.envsoft.2009.10.016

Qiu, H. (2012). *Study on the regional landslide characteristics analysis and hazard assessment: A case study of ningqiang county*. Xi'an: Northwest University.

Reichenbach, P., Busca, C., Mondini, A. C., and Rossi, M. (2014). The influence of land use change on landslide susceptibility zonation: The briga catchment test site (messina, Italy). *Environ. Manage.* 54 (6), 1372–1384. doi:10.1007/s00267-014-0357-0

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* 180, 60–91. doi:10.1016/j.earscirev.2018.03.001

Remondo, J., González, A., De Terán, J. R. D., Cendrero, A., Fabbri, A., and Chung, C. F. (2003). Validation of landslide susceptibility maps; examples and applications from a case study in northern Spain. *Nat. Hazards* 30, 437–449. doi:10.1023/B:NHAZ.0000007201.80743.fc

Rossi, M., Guzzetti, F., Rechenbach, P., Mondini, A. C., and Peruccacci, S. (2010). Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114, 129–142. doi:10.1016/j.geomorph.2009.06.020

Rowden, K. W., and Aly, M. H. (2018). A novel triggerless approach for mass wasting susceptibility modeling applied to the boston mountains of Arkansas, USA. *Nat. Hazards* 92 (1), 347–367. doi:10.1007/s11069-018-3201-7

Sabatakakis, N., Koukis, G., Vassiliades, E., and Lainas, S. (2013). Landslide susceptibility zonation in Greece. *Nat. Hazards* 65, 523–543. doi:10.1007/s11069-012-0381-4

Saha, S., Roy, J., Pradhan, B., and Hembram, T. K. (2021). Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at east Sikkim himalayan, India. *Adv. Space Res.* 68, 2819–2840. doi:10.1016/j.asr.2021.05.018

Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Aedizzone, F., et al. (2017). Characterization and quantification of path dependency in landslide susceptibility. *Geomorphology* 292, 16–24. doi:10.1016/j.geomorph.2017.04.039

San, B. T. (2014). An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The candir catchment area (western antalya, Turkey). *Int. J. Appl. Earth Obs.* 26, 399–412. doi:10.1016/j.jag.2013.09.010

Shao, X., Ma, S., Xu, C., and Zhou, Q. (2020). Effects of sampling intensity and non-slide/slide sample ratio on the occurrence probability of coseismic landslides. *Geomorphology* 363, 107222. doi:10.1016/j.geomorph.2020.107222

Soeters, R., and van Westen, C. J. (1996). "Slope instability recognition, analysis, and zonation," in *Landslides: Investigation and mitigation. Transportation research board special report 247*. Editors K. A. Turner and R. L. Schuster (National Research Council, Washington, USA).

Steger, S., Brenning, A., Bell, R., Petschko, H., and Glade, T. (2016). Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262, 8–23. doi:10.1016/j.geomorph.2016.03.015

Stokes, A., Atger, C., Bengough, A. G., Fourcaud, T., and Sidle, R. C. (2009). Desirable plant root traits for protecting natural and engineered slopes against landslides. *Plant Soil* 324 (1-2), 1–30. doi:10.1007/s11104-009-0159-y

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

Sun, X. (2020). *Study on landslide susceptibility and risk mapping along the rapidly uplifting section of the upper jinsha river: A case of xulong to benzilan reach*. Changchun: Jilin University. doi:10.27162/d.cnki.gjlin.2020.001120

Taner San, B. (2014). An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The candir catchment area (western antalya, Turkey). *Int. J. Appl. Earth Obs.* 26, 399–412. doi:10.1016/j.jag.2013.09.010

Trigila, A., Iadanza, C., Esposito, C., and Scarascia-Mugnozza, G. (2015). Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in giampilieri (NE sicily, Italy). *Geomorphology* 249, 119–136. doi:10.1016/j.geomorph.2015.06.001

van Westen, C. J., Castellanos, E., and Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* 102 (3–4), 112–131. doi:10.1016/j.enggeo.2008.03.010

Wu, W., and Sidle, R. C. (1995). A distributed slope stability model for steep forested basins. *Water Resour. Res.* 31, 2097–2110. doi:10.1029/95WR01136

Wu, X., Ren, F., Niu, R., and Peng, L. (2013). Landslide spatial prediction based on slope units and support vector machines. *Geomatics Inf. Sci. Wuhan Univ.* 38 (12), 1499–1503. doi:10.13203/j.whugis2013.12.006

Xu, C., and Xiao, J. Z. (2013). Spatial analysis of landslides triggered by the 2013 Ms7.0 lushan Earthquake: A case study of a typical rectangle area in the northeast of taiping town. *Seismol. Geol.* 35 (2), 436–451. doi:10.3969/j.issn.0253-4967.2013.02.021

Yalcin, A. (2008). GIS-Based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in ardesen (Turkey): Comparisons of results and confirmations. *Catena* 72, 1–12. doi:10.1016/j.catena.2007.01.003

Yan, F., Zhang, Q., Ye, S., and Ren, B. (2019). A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model. *Geomorphology* 327, 170–187. doi:10.1016/j.geomorph.2018.10.024

Yang, G., Xu, P., Cao, C., Zhang, W., Lan, Z., Chen, J., et al. (2019). Assessment of regional landslide susceptibility based on combined model of certainty factor method. *J. Eng. Geol.* 27 (5), 1153–1163. doi:10.13544/j.cnki.jeg.2019018

Yang, J., Song, C., Yang, Y., Xu, C., Guo, F., and Xie, L. (2019). New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of duwen highway basin, sichuan province, China. *Geomorphology* 324, 62–71. doi:10.1016/j.geomorph.2018.09.019

Yang, Z., Zhao, X., Chen, M., Zhang, J., Yang, Y., Chen, W., et al. (2023). Characteristics, dynamic analyses and hazard assessment of debris flows in niumiangou valley of wenchuan county. *Appl. Sci.* 13, 1161. doi:10.3390/app13021161

Yao, X., Deng, J., Liu, X., Zhou, Z., Yao, J., Dai, F., et al. (2020). Primary recognition of active landslides and development rule analysis for Pan three-river-parallel territory of tibet plateau. *Adv. Eng. Sci.* 52 (5), 16–37. doi:10.15961/j.jsuese.202000529

Yin, Y., Zhang, C., Yan, H., Xiao, M., Hou, X., Zhu, S., et al. (2022). Research on seepage stability and prevention design of landslides during impoundment operation of the Three Gorges Reservoir, China. *Chin. J. Rock Mech. Eng.* 41 (4), 649–659. doi:10.13722/j.cnki.jrme.2021.0674

Youssef, A. M., and Pourghasemi, H. R. (2021). Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at abha basin, asir region, Saudi arabia. *Geosci. Front.* 12, 639–655. doi:10.1016/j.gsf.2020.05.010

Yu, P. (2000). Application of physically-based distributed models in forest hydrology. *For. Res.* 4, 431–438. doi:10.13275/j.cnki.lykxyj.2000.04.015

Zêzere, J. L. (2002). Landslide susceptibility assessment considering landslide typology. A case study in the area north of lisbon (Portugal). *Nat. Hazards Earth Syst. Sci.* 2, 73–82. doi:10.5194/nhess-2-73-2002

Zhang, C., Yin, Y., Yan, H., Li, H., Dai, Z., and Zhang, N. (2021). Reactivation characteristics and hydrological inducing factors of a massive ancient landslide in the three gorges reservoir, China. *Eng. Geol.* 292 (1), 106273. doi:10.1016/j.enggeo.2021.106273

Zhang, Q., Ling, S., Li, X., Sun, C., Xu, J., and Huang, T. (2020). Comparison of landslide susceptibility mapping rapid assessment models in Jiuzhaigou County, Sichuan province, China. *Chin. J. Rock Mech. Eng.* 39 (8), 1595–1610. doi:10.13722/j.cnki.jrme.2020.0029

Zhang, Y., Qin, S., Zhai, J., Li, G., Peng, S., Liu, X., et al. (2018). Susceptibility assessment of debris flow based on GIS and weight information for the changbai mountain area. *Hydrogeology Eng. Geol.* 45 (2), 150–158. doi:10.16030/j.cnki.issn.1000-3665.2018.02.23

Zhang, Z., Deng, M., Xu, S., Zhang, Y., Fu, H., and Li, Z. (2022). Comparison of landslide susceptibility assessment models in zhenkang county, yunnan province, China. *Chin. J. Rock Mech. Eng.* 41 (1), 157–171. doi:10.13722/j.cnki.jrme.2021.0360

Zhang, Z. (2016). *Land use/land cover classification and change detection in mountainous areas using multi-resolution remotely sensed data – a case study of manwan reservoir basin*. Kunming: Yunnan University.

Zhao, F. (2020). *Early identification of geological hazards and landslides susceptibility evaluation of karakorum highway (domestic section)*. Lanzhou: Lanzhou University. doi:10.27204/d.cnki.glzhu.2020.000254

Zhao, X., Li, G., Zhao, Z., Li, C., Chen, Q., and Ye, X. (2022). Identifying the spatiotemporal characteristics of individual red bed landslides: A case study in western yunnan, China. *J. Mt. Sci.* 19 (6), 1748–1766. doi:10.1007/s11629-022-7339-0

Zhao, X., Zhang, H., Zhao, Z., Xia, L., Liu, B., Li, Y., et al. (2020). Study on the genesis of rainfall-glacier mixed type debris flow of Haibalo gully in northwest Yunnan on july 28, 2019. *J. Eng. Geol.* 28 (6), 1339–1349. doi:10.13544/j.cnki.jeg.2020-505

Zhao, Y., Wang, R., Jiang, Y., Liu, H., and Wei, Z. (2019). GIS-Based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in yueqing, southeastern China. *Eng. Geol.* 259, 105147. doi:10.1016/j.enggeo.2019.105147

Zheng, T., Cai, H., and Liu, J. (2022). Effects of land use and climate change on spatial distribution characteristics of hydrological factors in tongzi river basin [J/OL]. Journal of Yangtze River Scientific Research Institute. Available at: https://kns.cnki.net/kcms/detail/42.1171.TV.20220513.1529.005.html.

Zhu, A-X., Miao, Y., Yang, L., Bai, S., Liu, J., Hong, H., et al. (2018). Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. *Catena* 171, 222–233. doi:10.1016/j.catena.2018.07.012

Zhu, Q., Zhang, M., Ding, Y., Zeng, H., Wang, W., and Liu, F. (2021). Fuzzy logic approach for regional landslide susceptibility analysis constrained by spatial characteristics of environmental factors. *Geomatics Inf. Sci. Wuhan Univ.* 46 (10), 1431–1440. doi:10.13203/j.whugis20200653

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 30 March 2023
DOI 10.3389/feart.2023.1118160

OPEN ACCESS

EDITED BY
Wei Zhao,
Institute of Mountain Hazards and Environment (CAS), China

REVIEWED BY
Shuai Chen,
Central South University, China
Paraskevas Tsangaratos,
National Technical University of Athens, Greece
Himan Shahabi,
University of Kurdistan, Iran

*CORRESPONDENCE
Shengwu Qin,
qinsw@jlu.edu.cn

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 07 December 2022
ACCEPTED 21 March 2023
PUBLISHED 30 March 2023

CITATION
Lv J, Qin S, Chen J, Qiao S, Yao J, Zhao X, Cao R and Yin J (2023), Application of different watershed units to debris flow susceptibility mapping: A case study of Northeast China.
Front. Earth Sci. 11:1118160.
doi: 10.3389/feart.2023.1118160

Application of different watershed units to debris flow susceptibility mapping: A case study of Northeast China

Jiangfeng Lv[1,2], Shengwu Qin[1]*, Junjun Chen[1], Shuangshuang Qiao[1], Jingyu Yao[1], Xiaolan Zhao[2], Rongguo Cao[2] and Jinhang Yin[2]

[1]College of Construction Engineering, Jilin University, Changchun, China, [2]China Water Resources Bei Fang Investigation, Design & Research Co. LTD., Tianjin, China

The main purpose of this study was to compare two types of watershed units divided by the hydrological analysis method (HWUs) and mean curvature method (CWUs) for debris flow susceptibility mapping (DFSM) in Northeast China. Firstly, a debris flow inventory map consisting of 129 debris flows and 129 non-debris flows was randomly divided into a ratio of 70% and 30% for training and testing. Secondly, 13 influencing factors were selected and the correlations between these factors and the debris flows were determined by frequency ration analysis. Then, two types of watershed units (HWUs and CWUs) were divided and logistic regression (LR), multilayer perceptron (MLP), classification and regression tree (CART) and Bayesian network (BN) were selected as the evaluation models. Finally, the predictive capabilities of the models were verified using the predictive accuracy (ACC), the Kappa coefficient and the area under the receiver operating characteristic curve (AUC). The mean AUC, ACC and Kappa of four models (LR, MLP, CART and BN) in the training stage were 0.977, 0.931, and 0.861, respectively, for the HWUs, while 0.961, 0.905, and 0.810, respectively, for the CWUs; in the testing stage, were 0.904, 0.818, and 0.635, respectively, for the HWUs, while 0.883, 0.800, and 0.601, respectively, for the CWUs, which showed that HWU model has a higher debris flow prediction performance compared with the CWU model. The CWU-based model can reflect the spatial distribution probability of debris flows in the study area overall and can be used as an alternative model.

KEYWORDS

debris flow susceptibility mapping, watershed units, hydrological analysis method, mean curvature method, machine learning model

1 Introduction

According to the China Statistical Yearbook (http://www.stats.gov.cn/tjsj/ndsj/), a total of 7,840 geological disasters occurred in China in 2020, resulting in 197 casualties and direct economic losses of 740 million dollars, of which debris flows accounted for 11.46%. Debris flows are among the most frequent and destructive disasters in mountainous areas (Dash et al., 2022; Jiang et al., 2022; Qiu et al., 2022). Debris flow susceptibility mapping (DFSM), representing where debris flows are likely to occur, plays an important role in debris flow management strategies and has been a hot

topic in disaster research worldwide (Ilia and Tsangaratos, 2015; Qin et al., 2019; Sun et al., 2021; Yao et al., 2022).

There are many uncertainties in the process of disaster susceptibility mapping, such as selecting appropriate mapping units, determining evaluation models, screening influencing factors, determining the proportion of training and testing data and others (Tien Bui et al., 2015; Cama et al., 2016; Zezere et al., 2017; Chen et al., 2018; Du et al., 2018; Dou et al., 2019; Qiao et al., 2021). Among the above uncertainty factors, selecting appropriate mapping units is the first step to address disasters and environmental factors. The mapping unit is the basic functional spatial element for dividing the study area (Cama et al., 2016). The term refers to a portion of the land surface which contains a set of ground conditions that differ from the adjacent units across definable boundaries (Van Den Eeckhaut et al., 2009). The selection of mapping units affects the methods used to address the uncertainty in the input data, the model fitting, the reliability of disaster susceptibility mapping and the application of disaster susceptibility mapping in disaster prevention and mitigation (Fausto Guzzetti et al., 1999; Cama et al., 2016; Qiao et al., 2021). At present, mapping units mainly include the following classes: grid cell units, slope units, watershed units, topographic units, geohydrological units, political or administrative units, and unique condition units (Van Den Eeckhaut et al., 2009; Chen et al., 2019; Sun et al., 2020).

For DFSM, grid cell units and watershed units are used frequently. Grid cell units are the most popular mapping units

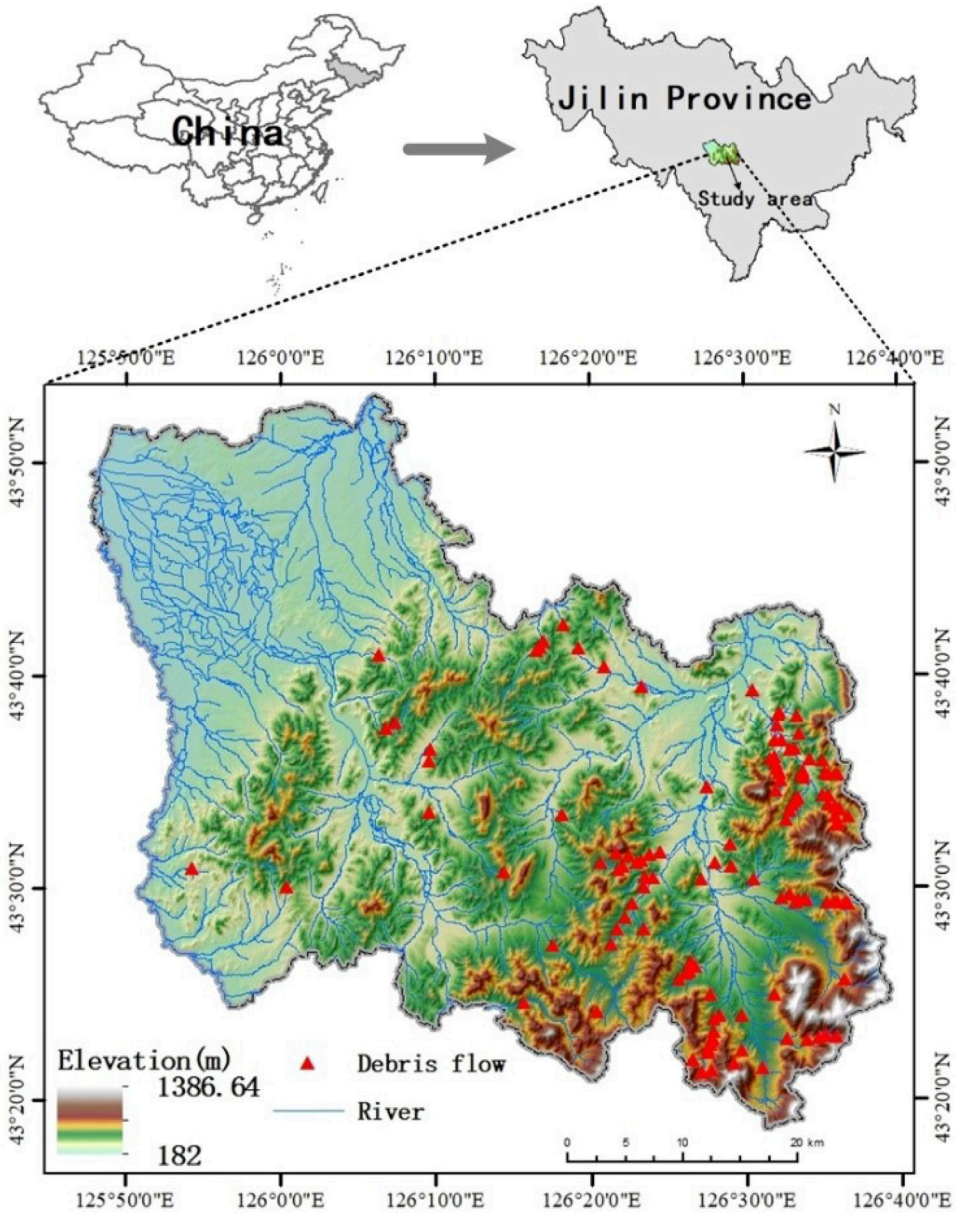


FIGURE 1
The geographic location of the study area.

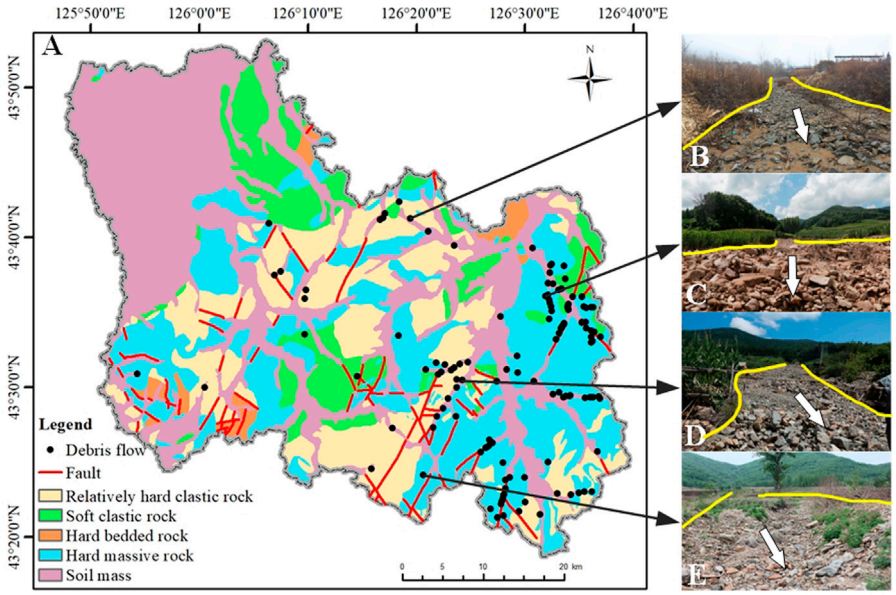


FIGURE 2
Geological map and debris flow field photos of the study area: **(A)** geological map; **(B–E)** debris flow field photos.

with the same cell size, fast processing speed and simple algorithm (Reichenbach et al., 2018). However, the division of grid cells destroys the integrity of debris flows and is almost completely unrelated to geological and topographic information (Dragut and Eisank, 2011; Wang et al., 2017). Moreover, since debris flows are a dynamic process, the DFSM based on grid cell units cannot comprehensively reflect spatial information (Qin et al., 2019). Watershed refers to the river catchment area that is surrounded by the water-parting line; it is the basic unit for the development and activity of debris flows, and it is the object of exploration, research, and prevention of debris flows. Furthermore, the watershed unit includes the formation area, circulation area, and accumulation area of a debris flow (Qin et al., 2019). Compared with grid cell units, watershed units can completely consider the spatial information of a debris flow. Some scholars have carried out DFSM based on watershed units and obtained reliable results. Qin et al. (2019) explored the accuracy and practicability of mapping units for the evaluation of debris flow susceptibility based on grid cell units and watershed units, and the results showed that watershed units were more feasible than grid cell units when considering the effects of geology and geomorphology on the occurrence of debris flows. Qiao et al. (2021) proposed a region-partitioning method for DFSM based on the topographic characteristics of watershed units, and the results demonstrated that this method can enable more reasonable regional-scale DFSM. Li et al. (2017) presented an application of the rock engineering system and fuzzy C-means algorithm for debris flow susceptibility assessment using watershed units as mapping units in the Wudongde Dam area, the evaluation results agreed well with field investigations. Zou et al. (2019) developed a quantitative method for regional risk assessment of debris flows by analyzing in-depth the relationships among hazard-forming environments, disaster factors and elements at risk based on hydrological response units. The presented method may serve as pertinent guidance for regional risk assessment of debris flows. In addition, some scholars have used watershed units to evaluate and compare the performance of different evaluation models for DFSM (Liang et al., 2020; Xiong et al., 2020), and the conclusions provide helpful data for assessing and mitigating debris flow hazards. Therefore, it is important to carry out research based on watershed units, which provide more evidence and views for DFSM research. The commonly used watershed units are based on the hydrological analysis model, also known as hydrological response units (Li et al., 2021). In addition, watershed units can be generated based on the mean curvature model (Romstad and Etzelmüller, 2012). To compare the results of applying different watershed units in DFSM, we extracted the watershed units based on the hydrological analysis method and mean curvature method in the study.

There are plenty of evaluation models for disaster susceptibility mapping, from qualitative approaches to quantitative approaches (Aditian et al., 2018; Huang et al., 2020; Asadi et al., 2022). Qualitative methods are based on air photo and field interpretation and the opinions of an individual or a group of experts (Aditian et al., 2018; Ghasemian et al., 2022b). Some qualitative methods include ranking and weighting, such as analytic hierarchy process and weighted linear combination (Ayalew and Yamagishi 2005; Rozos et al., 2010). These qualitative or semi-quantitative methods are subjective and highly dependent on experts' knowledge, and are not suitable for large-scale research fields (Bălteanu et al.,

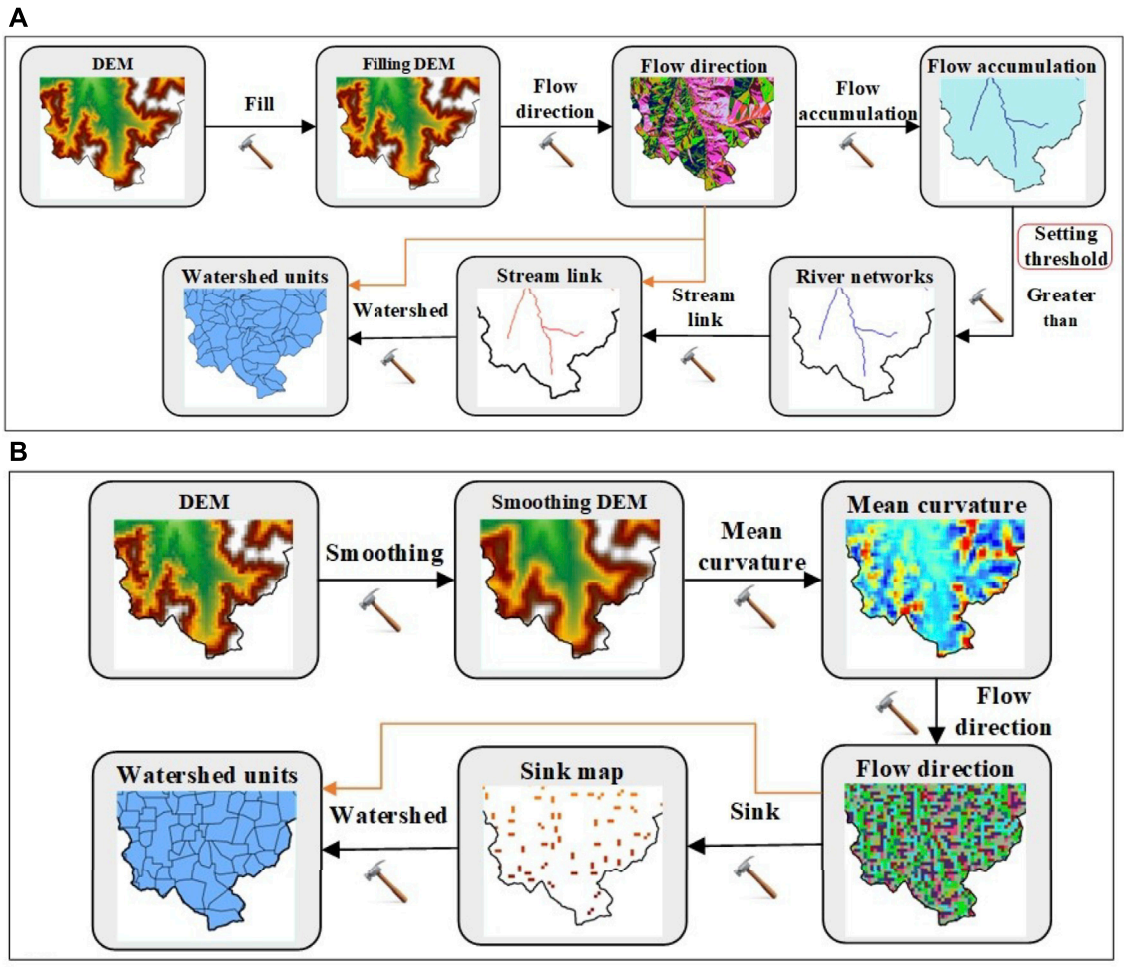


FIGURE 3
Classification process of the watershed units: **(A)** hydrological analysis method and **(B)** mean curvature method.

2010). Quantitative statistical models are built based on appropriate mathematical models to analyze the statistical relations between disasters and influencing factors (Hadmoko et al., 2017; Ghasemian et al., 2022b), including the information value (Xu et al., 2012), certainty factor concepts (Devkota et al., 2012), frequency ratio method (Balamurugan et al., 2016), bivariate statistical analysis (Ayalew and Yamagishi 2005), index of entropy (Shirani et al., 2018), weight of evidence (Constantin et al., 2010), evidential belief functions (Carranza 2014), logistic regression (Cao et al., 2019), *etc.* Machine learning models are now widely used because these models can analyze the non-linear corrections between past events and the influencing factors and they predict where disasters will occur (He et al., 2012; Xiong et al., 2020). These models include artificial neural networks (Pham et al., 2017; Chen et al., 2021; Chen et al., 2022), support vector machines (Colkesen et al., 2016), random forest (Hong et al., 2016), decision trees (Althuwaynee et al., 2014), classification and regression tree (Youssef et al., 2015), boosted regression trees (Xiong et al., 2020), Bayesian network (Song et al., 2012), adaptive neuro-fuzzy inference (Jaafari et al., 2019), logistic model tree (Tien Bui et al., 2015) and random gradient descent (Hong et al., 2020). Reichenbach et al. (2018) reviewed the statistically-based landslide susceptibility assessment literature from 1983 to 2016, and found that the most common statistical methods for landslide susceptibility modeling include logistic regression, neural network analysis, data-overlay and index-based and weight of evidence analyses. In this study, to avoid the model uncertainty caused by different evaluation models, we use logistic regression (LR), multilayer perceptron (MLP), classification and regression tree (CART) and Bayesian network (BN) to carry out DFSM based on two types of watershed units.

This study compared and analysed the applicability of two different watershed units in regional DFSM based on four models (LR, MLP, CART, and BN). The main purpose is to support the selection of watershed units for DFSM. Yongji county in the Jilin Province, China was taken as the study region because it is under serious threat of frequent debris flows. The division process and results of two types of watershed units were compared. Eight DFSMs are discussed and AUC, ACC, and Kappa analyses were used to evaluate the accuracy of the debris flow susceptibility models.

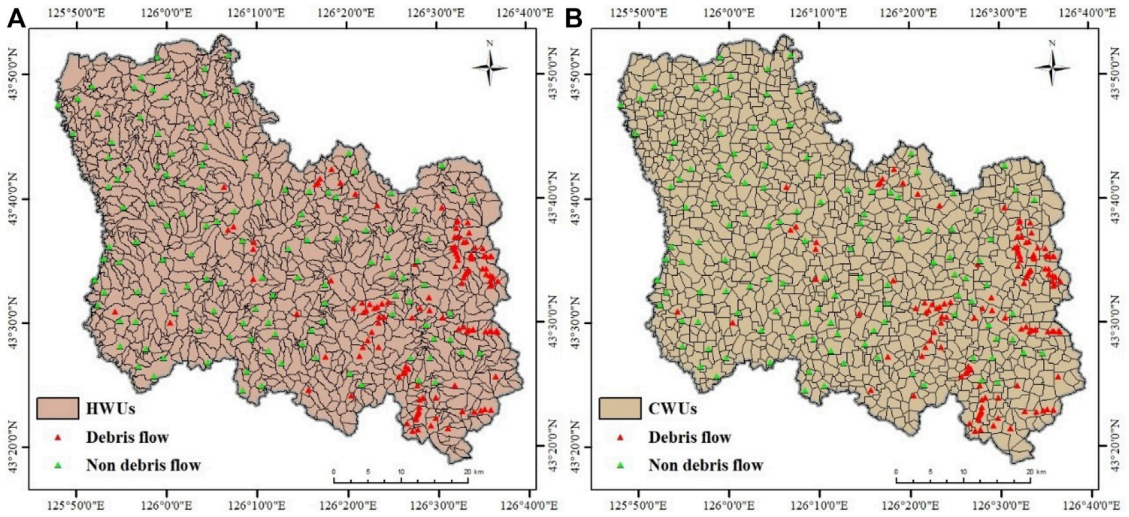


FIGURE 4
Division of watershed units: **(A)** hydrological analysis method and **(B)** mean curvature method.

2 Study area

2.1 General settings

Yongji county is located in central eastern Jilin Province, China (Figure 1), which covers a total area of 2,620 km^2. The number of debris flows in Yongji county has increased from 71 in 2007 to 129 in 2021, causing several deaths, destroying hundreds of houses and thousands of acres of farmland. The debris flows scoured the roadbed and piled up on the road, resulting in traffic paralysis. It is necessary and urgent to map the susceptibility of debris flows in Yongji county.

The study area lies between 125°48′09″E to 126°40′01″E longitude and 43°18′07″N to 43°35′00″N latitude. There are four landforms in the entire area: middle mountains, low mountains, platform, and river valley. From southeast to northwest, the landforms of the study area are middle mountains, low mountains and platform with the altitudes ranging from 1,386 to 182 m. In addition to several andesites and metamorphic rocks, the main rock type is Yanshan Early Granite. The study area lies in the Tianshan–Xingan geosyncline fold area of the Jilin and Heilongjiang fold system (Qin et al., 2019). Folds and faults are relatively developed in Yongji county, which provides conditions for the occurrence of geological disasters (Figure 2A). Yongji county is in the mid-latitude subtemperate continental climate zone with an annual average precipitation of 722.75 mm. There are 39 rivers covering an area of more than 20 km^2. The main rivers include the Yinma River, Wende River, Chalu River and Aolong River.

2.2 Debris flow data inventory

A debris flow inventory map is a prerequisite for DFSM(Xu et al., 2012; Arabameri et al., 2020; Dash et al., 2022). A total of 129 debris flows were collected based on field surveys and historical materials. Figure 2A shows that debris flows are mainly distributed across the southeast mountain area. Statistics show that among 129 debris flows, only 7 are medium in size and 122 are small. In recent years, the increase in debris flow frequency in Yongji county has been closely related to deforestation and reclamation. With the destruction of forest vegetation, rainfall is more likely to cause soil erosion, which gradually forms a series of gullies. These gullies provide circulation conditions for debris flows. Figures 2B–E shows some images of occurred debris flows in the study area.

3 Watershed units

3.1 Division methods for different watershed units

In this study, the extraction of watershed units was completed in ArcGIS 10.2 software (Tien Bui et al., 2015; Cao et al., 2019). The most commonly used watershed units (HWUs) are classified by the hydrological principles (Fausto Guzzetti et al., 1999). HWUs are derived based on an 8-direction flow algorithm (Horton et al., 2013). Establishing the HWUs consists of the following six steps: 1) filling the original DEM, 2) extracting the flow direction, 3) calculating the flow accumulation, 4) extracting river networks based on a threshold, 5) stream linking, 6) dividing HWUs based on flow direction and stream linking. The detailed classification process is shown in Figure 3A.

In addition, watershed units can be generated based on the mean curvature method (CWUs). The mean curvature is a simple combination of profile curvature and plan curvature. Its maximum and minimum values can indicate the changes in aspect and slope positions at the same time. Therefore, the mean curvature can reflect the ridge line, valley line, platform edge and wide valley edge (Romstad and Etzelmüller, 2012). Establishing the CWUs consists of the following five steps: 1) smoothing the original

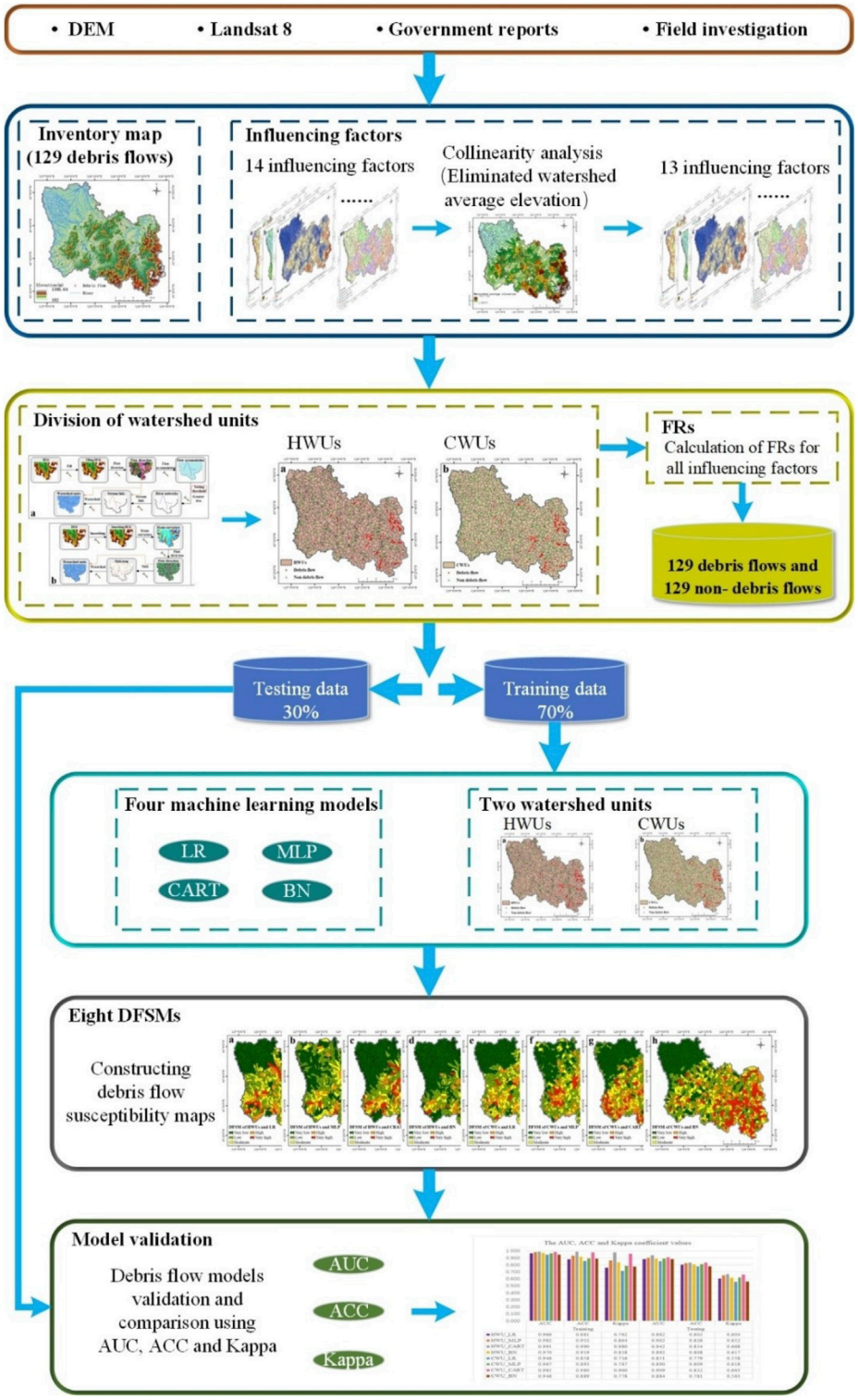


FIGURE 5
Flowchart of the research methodology.

DEM, 2) calculating the mean curvature, 3) extracting the flow direction, 4) filling depressions based on flow direction data, and 5) dividing CWUs based on flow direction and depressions. The detailed classification process is shown in Figure 3B.

3.2 Watershed unit classification results

For HWUs, the number and size are closely related to DEM resolution and flow threshold, but for CWUs, the control factor is

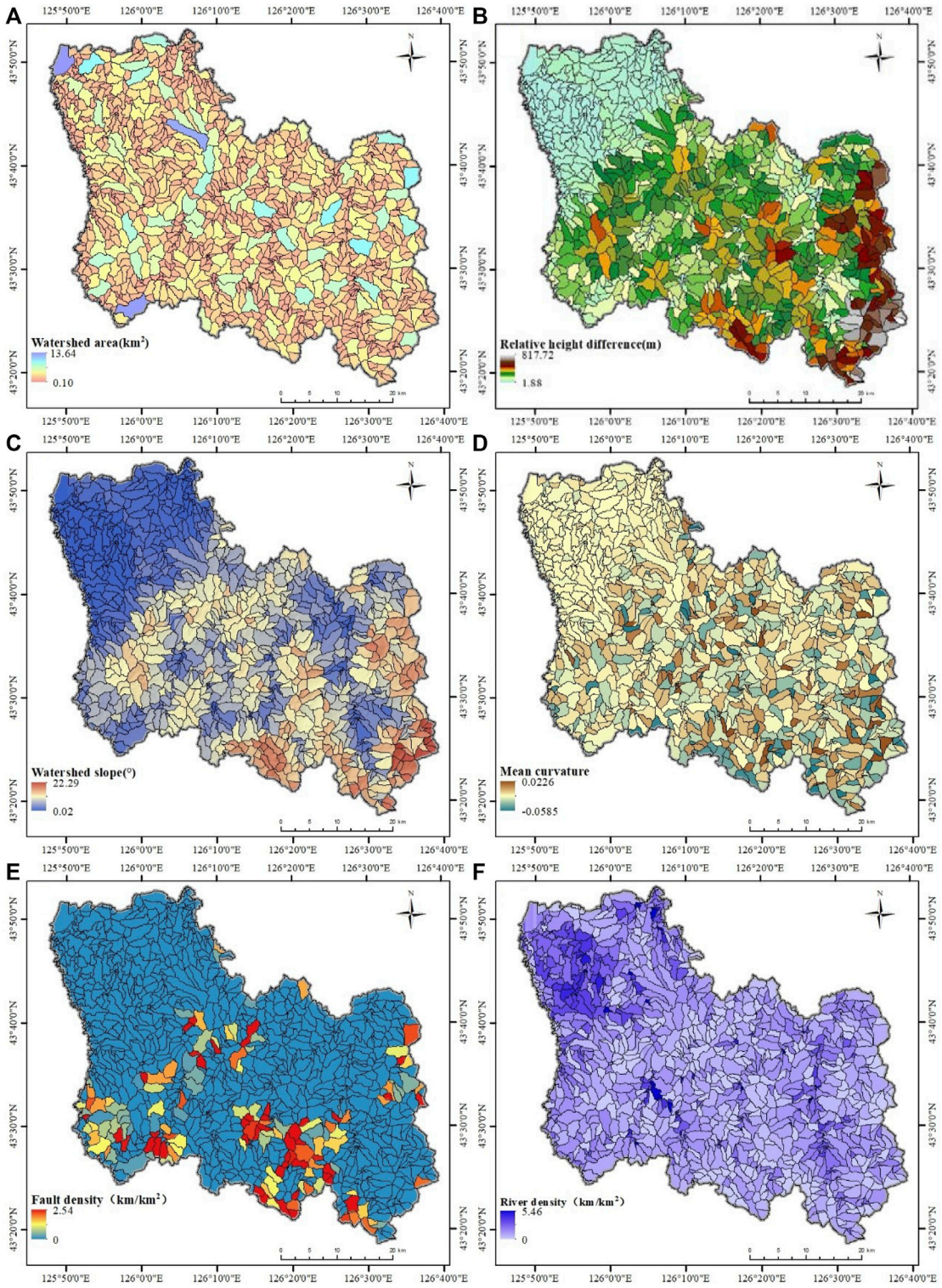


FIGURE 6
(Continued).

only DEM resolution. For HWUs, flow threshold values of 500, 1,000, 2000, 5,000, and 10,000 were chosen based on a DEM with a resolution of 30 m. For CWUs, we resampled the DEM with resolutions of 50, 100, 200, 300, 500, and 1,000. To ensure that the number and size of the two types of watershed units were not much different and consistent with the actual watersheds, a flow

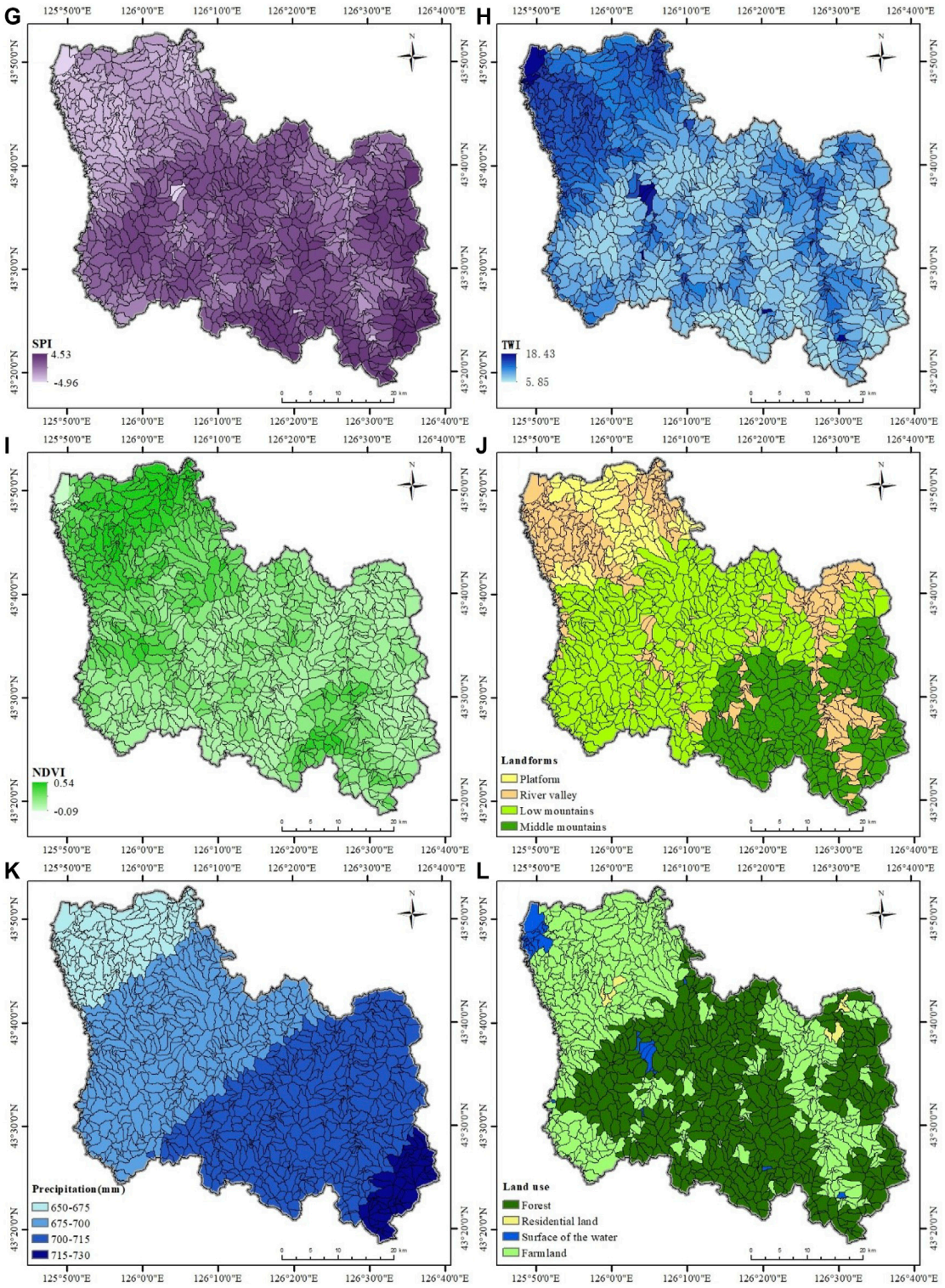


FIGURE 6
(Continued).

threshold of 1,000 and a DEM resolution of 300 were selected to divide the watershed units. For the HWUs, the study area was divided into 1,092 watershed units. The smallest unit was 0.10 km^2, the largest unit was 13.63 km^2, and the mean size was 2.40 km^2 (Figure 4A). For CWUs, the study area was divided into 1,211 watershed units. The smallest unit was 0.11 km^2, the largest unit was 8.87 km^2, and the mean size was 2.17 km^2 (Figure 4B).

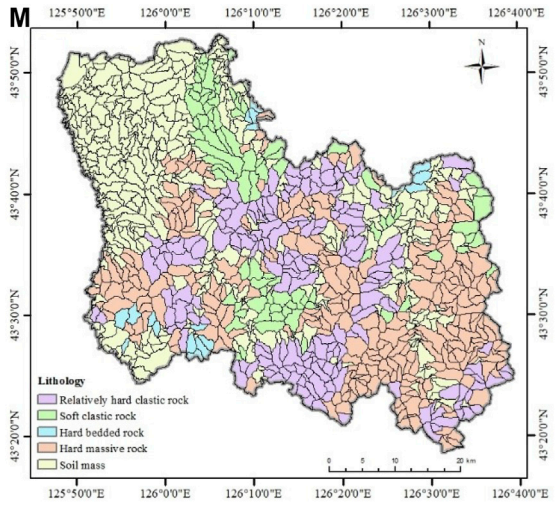


FIGURE 6
(Continued). Maps of influencing factors based on HWUs with a flow threshold of 1,000: **(A)** watershed area; **(B)** relative height difference; **(C)** watershed slope; **(D)** mean curvature; **(E)** fault density; **(F)** river density; **(G)** SPI; **(H)** TWI; **(I)** NDVI; **(J)** landforms; **(K)** precipitation; **(L)** land use; **(M)** lithology.

4 Materials and methods

The flowchart of the research methodology is shown in Figure 5. DFSM of Yongji county using four models (LR, MLP, CART, and BN) and watershed units (HWUs and CWUs) have been carried out in five main steps: 1) data collection and screening influencing factors, 2) division of two types of watershed units, 3) calculation of frequency ratio values (FRs) for all influencing factors, 4) building debris flow models and constructing DFSM, and 5) debris flow model validation and comparison using AUC, ACC and Kappa.

4.1 Screening influencing factors

The occurrence of debris flows is affected by many factors including topographic, geomorphologic, geological, ecological and meteorological factors (Zhang et al., 2012; Bregoli et al., 2014; Hu et al., 2014). Based on field observations, available literature and expert experience, fourteen influencing factors were considered, such as watershed area, relative height difference, watershed average elevation, watershed slope, mean curvature, fault density, river density, stream power index (SPI), topographic wetness index (TWI), plan normalized difference vegetation index (NDVI), landforms, precipitation, land use and lithology.

Because substantial collinearity will lead to model instability, collinearity analysis is essential before influencing factors are applied for DFSM(Qiu et al., 2022). Person's correlation coefficient was calculated to test the collinear relationship among these factors, and the results are shown in Table 1. There is no correlation coefficient when the absolute value is less than 0.7 (Dormann et al., 2013; Yao et al., 2022). There was high collinearity between relative height difference and watershed average elevation, and the Person's correlation coefficient was 0.86. In addition, the value of collinearity between watershed average elevation and precipitation was 0.69. Therefore, the watershed average elevation was eliminated.

The watershed area, relative height difference, watershed slope, mean curvature, SPI and TWI were extracted from the DEM with a resolution of 30 m. Fault, river, and lithology data were acquired from the geological map of Yongji county and field investigations. The Landsat 8 image taken on 11 August 2021, was used to produce the NDVI. Landforms, precipitation, and land use were provided by government reports. Thirteen influencing factors were converted to a grid cell with a resolution of 30 m in ArcGIS 10.2 (Chen et al., 2017). Table 2 shows date source and scale of influencing factors. When watershed units are applied to DFSM, grid patterns for each factor need to be transferred to the corresponding watershed units. For watershed area, geometric calculation in the attribute table was used to calculate the area of each watershed. The difference between the highest and the lowest points in each watershed was calculated as a relative height difference (Qin et al., 2019). For watershed slope, mean curvature, SPI, TWI, and NDVI, the zonal statistics tool in the spatial analysis was used and the statistical type was "mean." The length of faults and rivers in each watershed was extracted by using the intersection tool, and then, the fault density and river density in each watershed were calculated using the field calculator. Precipitation for each watershed was determined based on the principle of majority, and this principle was also applied to factors of landforms, land use and lithology. The data types of precipitation, landforms, land use and lithology are discrete, while the data types of other factors are continuous. The influencing factor layers based on HWUs with a flow threshold of 1,000 are shown in Figure 6.

The FRs of the influencing factor subclass were used as the input variable of the DFSM models (Huang et al., 2020). Based on a series of previous studies (Xu et al., 2012; Aditian et al., 2018; Vakhshoori et al., 2019; Chang et al., 2020), we divided the continuous factor into eight levels using the natural fracture method. Taking HWUs with a flow threshold of 1,000 as an example, the FRs for each level of thirteen factors are shown in Table 3.

4.2 Logistic regression (LR)

Logistic regression (LR) may be the most widely used statistical technique in susceptibility assessment (Colkesen et al., 2016). As a multivariate regression method, LR can find a model to describe the relationship between multiple independent variables and a dependent variable (Lee and Pradhan 2006; Lee 2007; Pourghasemi et al., 2013). For DFSM, the influencing factors are considered the independent variables and the occurrence and non-occurrence of debris flows are considered the dependent variables. For LR, variables may be continuous, discrete or arbitrary combinations of two types (Lee, 2007). LR can be expressed as follows (Ayalew and Yamagishi 2005; Yalcin et al., 2011; Schlögel et al., 2018):

$$P = \frac{1}{1 + e^{-z}} \tag{1}$$

$$Z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n \tag{2}$$

where P denotes the probability of a debris flow occurrence in each watershed, varying between 0 and 1; Z represents the dependent variable including non-debris flows (0) and debris flows (1); α represents the

TABLE 1 The results of the Person's Correlation Coefficient.

Factors	WA	E	RHD	WS	F	R	NDVI	SPI	TWI	Pre	MC	LU	LF	Li
WA	1.00													
E	0.04	1.00												
RHD	0.25	0.86	1.00											
WS	0.00	0.33	0.29	1.00										
F	−0.03	0.22	0.09	0.02	1.00									
R	−0.05	−0.44	−0.36	−0.36	−0.18	1.00								
NDVI	−0.10	−0.30	−0.38	−0.22	−0.08	0.17	1.00							
SPI	0.06	0.45	0.46	0.23	0.14	−0.19	−0.23	1.00						
TWI	0.00	0.00	0.04	−0.55	0.00	0.20	0.02	0.36	1.00					
Pre	−0.02	0.69	0.48	0.21	0.18	−0.24	−0.43	0.42	0.02	1.00				
MC	−0.10	−0.19	−0.19	−0.03	−0.08	0.03	0.06	−0.42	−0.25	−0.11	1.00			
LU	0.02	−0.17	−0.17	−0.59	0.11	0.31	0.18	−0.04	0.37	−0.08	−0.13	1.00		
LF	0.03	0.33	0.30	0.42	0.12	−0.34	−0.28	0.27	−0.16	0.36	−0.16	−0.26	1.00	
Li	0.00	−0.10	−0.09	−0.48	−0.04	0.23	0.13	−0.15	0.29	−0.03	0.05	0.31	−0.28	1.00

('WA' represents 'Watershed area', 'E' represents 'Watershed average elevation', 'RHD' represents 'Relative height difference', 'WS' represents 'Watershed slope', 'F' represents 'Fault density', 'R' represents 'River density', 'NDVI' represents 'Plan normalized difference vegetation index', 'SPI' represents 'Stream power index', 'TWI' represents 'topographic wetness index', 'Pre' represents 'precipitation', 'MC' represents 'mean curvature', 'LU' represents 'Land use', 'LF' represents 'Landforms' and 'Li' represents 'Lithology').

TABLE 2 Date source and scale of influencing factors.

Factors	Data source	Scale
Watershed area (km^2)	DEM	30 m × 30 m
Relative height difference (m)	DEM	30 m × 30 m
Watershed slope (°)	DEM	30 m × 30 m
Mean curvature	DEM	30 m × 30 m
Fault density (km/km^2)	The geological map of Yongji County and field investigations	1:200000
River density (km/km^2)	The geological map of Yongji County and field investigations	1:200000
SPI	DEM	30 m × 30 m
TWI	DEM	30 m × 30 m
NDVI	The Landsat 8 image	30 m × 30 m
Landforms	Government reports	1:200000
Precipitation (mm)	Government reports	1:200000
Land use	Government reports	1:200000
Lithology	The geological map of Yongji County and field investigations	1:200000

intercept of the regression function, β_1, β_2, … β_n are the regression coefficients; and x_1, x_2, … x_n are the debris flow influencing factors.

4.3 Multilayer perceptron (MLP)

Multilayer perceptron (MLP) is a kind of artificial neural network and has been widely used in classification (Tien Bui et al., 2015; Pham et al., 2017). The MLP generally consists of three main components, namely, input layers, hidden layers, and output layers (Kavzoglu and Mather 2003). For DFSM, the input layers are considered the influencing factors of debris flow, the output layers are considered the classification result of inferring debris flow or non-debris flow, and the hidden layers are considered the classification layers that convert input into output. The MLP model with only one hidden layer is the most basic three-tier structure model, which can fit and predict many

TABLE 3 Description and FRs of all the influencing factors (HWUs with a flow threshold of 1,000).

Factors	Descriptions of influencing factors	Class	FR
Watershed area (km²)	The watershed area is generally proportional to the amount of water in the catchment and material sources. The greater the material sources are, the stronger the destructive power is	0.10–1.17	0.92
		1.17–2.01	0.36
		2.01–2.84	1.01
		2.84–3.86	1.43
		3.86–5.03	0.95
		5.03–6.64	0.14
		6.64–8.84	3.14
		8.94–13.64	1.77
Relative height difference (m)	The relative elevation difference is the value between the highest and lowest elevations in a watershed unit Shi et al. (2015). The greater the height difference is, the greater the potential energy of the debris flow is, loose deposits are easy to move under water scouring, and the kinetic energy of debris flow is also high	1.88–58.41	0.00
		58.41–125.09	0.22
		125.09–187.63	0.82
		187.63–251.52	0.80
		251.52–323.73	1.28
		323.73–429.99	1.76
		429.99–590.74	3.95
		590.74–817.72	2.98
Watershed slope (°)	The watershed slope represents the average slope within the watershed. The greater the watershed slope is, the worse the slope stability is, which provides materials for debris flows	0.02–3.77	0.00
		3.77–5.54	0.68
		5.54–7.42	0.80
		7.42–9.40	1.25
		9.40–11.46	2.18
		11.46–14.46	2.29
		14.46–22.29	2.43
Mean curvature	Curvature is generally used to describe the physical characteristics of the watershed and understand the erosion process Li et al. (2021)	−0.0585–−0.0094	2.34
		−0.0094–−0.0037	2.15
		−0.0037–−0.0016	1.93
		−0.0016–−0.0002	1.18
		−0.0002–0.0009	0.50

(Continued on following page)

10.3389/feart.2023.1118160

TABLE 3 (*Continued*) Description and FRs of all the influencing factors (HWUs with a flow threshold of 1,000).

Factors	Descriptions of influencing factors	Class	FR
		0.0009–0.0028	0.87
		0.0028–0.0060	0.61
		0.0060–0.0226	4.84
Fault density (km/km^2)	Faults are often related to seismic activity. Earthquakes produce loose deposits which provides materials for debris flows Hong et al. (2015)	0–0.07	0.82
		0.07–0.19	3.22
		0.19–0.33	0.62
		0.33–0.44	1.82
		0.44–0.58	1.45
		0.58–0.79	0.37
		0.79–1.11	4.41
		1.11–2.54	1.04
River density (km/km^2)	River density is the ratio of total river length in a watershed to the watershed area Lei et al. (2010). The river will erode the rock mass and destroy the stability of the slope, which provides conditions for the formation of debris flows	0–0.18	0.61
		0.18–0.46	0.78
		0.46–0.69	2.07
		0.69–0.97	1.10
		0.97–1.36	0.65
		1.36–1.90	0.22
		1.90–2.81	0.15
		2.81–5.46	0.00
SPI	SPI is used to measure the erosion power of the stream Althuwaynee et al. (2014)	−4.96–1.13	0.00
		1.13–1.73	0.14
		1.73–2.29	0.57
		2.29–2.81	0.87
		2.81–3.33	1.93
		3.33–4.53	2.50
TWI	The higher the TWI value is, the higher the soil water content is, indicating a higher potential for triggering debris flows in the watershed (Esper Angillieri. (2020)	5.85–6.79	2.15
		6.79–7.23	1.59
		7.23–7.73	1.03

(Continued on following page)

10.3389/feart.2023.1118160

TABLE 3 (*Continued*) Description and FRs of all the influencing factors (HWUs with a flow threshold of 1,000).

Factors	Descriptions of influencing factors	Class	FR
		7.73–8.32	0.54
		8.32–8.96	0.12
		8.96–18.43	0.00
NDVI	The NDVI reflects vegetation coverage. High vegetation coverage is conducive to soil and water conservation and can reduce the amount of materials in debris flows	−0.09–0.08	1.52
		0.08–0.12	1.31
		0.12–0.17	1.17
		0.17–0.23	1.53
		0.23–0.29	0.40
		0.29–0.36	0.99
		0.36–0.43	0.35
		0.43–0.54	0.21
Landforms	Landforms affect the formation, movement and scale of debris flows. Mountain areas are prone to debris flows due to the large slope and fast water catchment speed	Platform	0.00
		River valley	0.27
		Low mountains	0.37
		Middle mountains	3.15
Precipitation (mm)	Precipitation provides water and dynamic conditions for debris flows	650–675	0.00
		675–700	0.25
		700–715	2.05
		715–730	0.83
Land use	Land use is closely related to the occurrence of debris flows. For example, farmland is prone to debris flows due to serious soil erosion	Forest	1.45
		Residential land	1.04
		Surface of the water	0.00
		Farmland	0.51
Lithology	For different lithologies, their hardness, resistance to erosion and weathering are also different	Relatively hard clastic rock	0.94
		Soft clastic rock	0.87
		Hard bedded rock	0.00
		Hard massive rock	2.00
		Soil mass	0.25

10.3389/feart.2023.1118160

TABLE 4 Confusion matrix.

Actuality \ Prediction	Debris-flow (1)	Non-Debris-flow (0)
Debris-Flow (1)	True positive (TP)	False negative (FN)
Non-Debris-Flow (0)	False positive (FP)	True negative (TN)

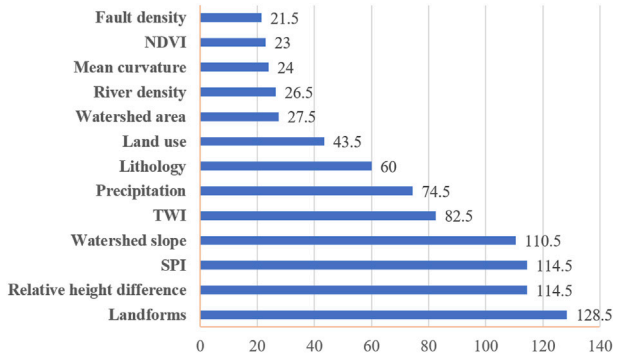


FIGURE 7
Contribution of influencing factors.

non-linear problems (Li et al., 2019). In this study, a single-hidden-layer MLP model is used in DFSM. For example, n_0, n_1 and n_2 represent the number of input, hidden and output layers, respectively, and the input variables are $X = [x_1, x_2 \ldots, x_{n0}]$. Then, the input and output of the hidden layer are: (Li et al., 2019; Huang et al., 2020):

$$z_j = \sum_{i=1}^{n_0} w_{ij} x_i + b_j \tag{3}$$

$$y_j = f(z_j) = \left(1 + e^{-zj}\right)^{-1} \tag{4}$$

where z_j, b_j and y_j represent the y th input variable, threshold value and output variable of the neuron in the hidden layer, respectively, w_{ij} represents the weight value between the i th input neuron and the j th neuron in the hidden layer, and $f(z_j)$ represents the activation function. Then the input and output of neurons in the output layer are:

$$z_k = \sum_{j=1}^{n_1} w_{jk} y_i + b_k \tag{5}$$

$$y_k = z_k \tag{6}$$

where z_k, b_k and y_k represent the j th input variable, threshold value and output variable of the neuron in the output layer, respectively, w_{ik} represents the weight value between the j th neuron in the hidden layer and the k th output neuron.

4.4 Classification and regression tree (CART)

The decision tree model is a technique that uses a tree structure to discover and describe structural patterns in data. It does not require a preestablished relationship between all input variables and a target variable (Hitoshi Saito and Matsuyama, 2009). As an algorithm of the decision tree model, classification and regression tree (CART) was first proposed by Breiman et al. (1984) The CART consists of a root node, a set of internal nodes and a set of leaf nodes. The leaf nodes correspond to the classification result, and the other nodes correspond to the classification rules. CART was selected as the decision tree model in this study in view of its performance efficiency (Wang et al., 2015).

4.5 Bayesian network (BN)

The Bayesian network (BN) is a graphical model for probabilistic relationships among a set of variables (Song et al., 2012). BN can be represented by directed acyclic graphs and conditional probabilities, reflecting the independent and interdependent relationship among various variables. The calculation formula is given as follows (Han et al., 2019):

$$P(L, M, N) = P(L) \times P(M|L) \times P(N|L, M) \tag{7}$$

where $P(L)$ is the prior probability, indicating the conditional probability without the parent node, $P(M|L)$ is the conditional probability, indicating the occurrence probability of M under L conditions and $P(N|L, M)$ is the conditional probability, indicating the occurrence probability of N under L and M conditions.

4.6 Model performance evaluation

In this study, three commonly used criteria, including the predictive accuracy (ACC), the Kappa coefficient and the area under the receiver operating characteristic curve (AUC) were used to evaluate the prediction ability of DFSMs. The calculation of the three criteria is based on the confusion matrix (Ghasemian et al., 2022a). The confusion matrix, also known as the error matrix, is a standard format for accuracy evaluation. The confusion matrix can represent the difference between the model prediction results and the actual observation results (Xiong et al., 2020). In this study, the confusion matrix of the debris flow susceptibility predictive models is shown in Table 4. For example, a true positive (TP) suggests that the prediction result is 'Debris-Flow', and the actual observation result is 'Debris-Flow'.

The predictive accuracy (ACC) represents the ratio of correctly predicted observations to total observations. This index shows how well the debris flow model works:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}$$

The Kappa index is used to assess the acceptability of debris flow models which can be calculated by:

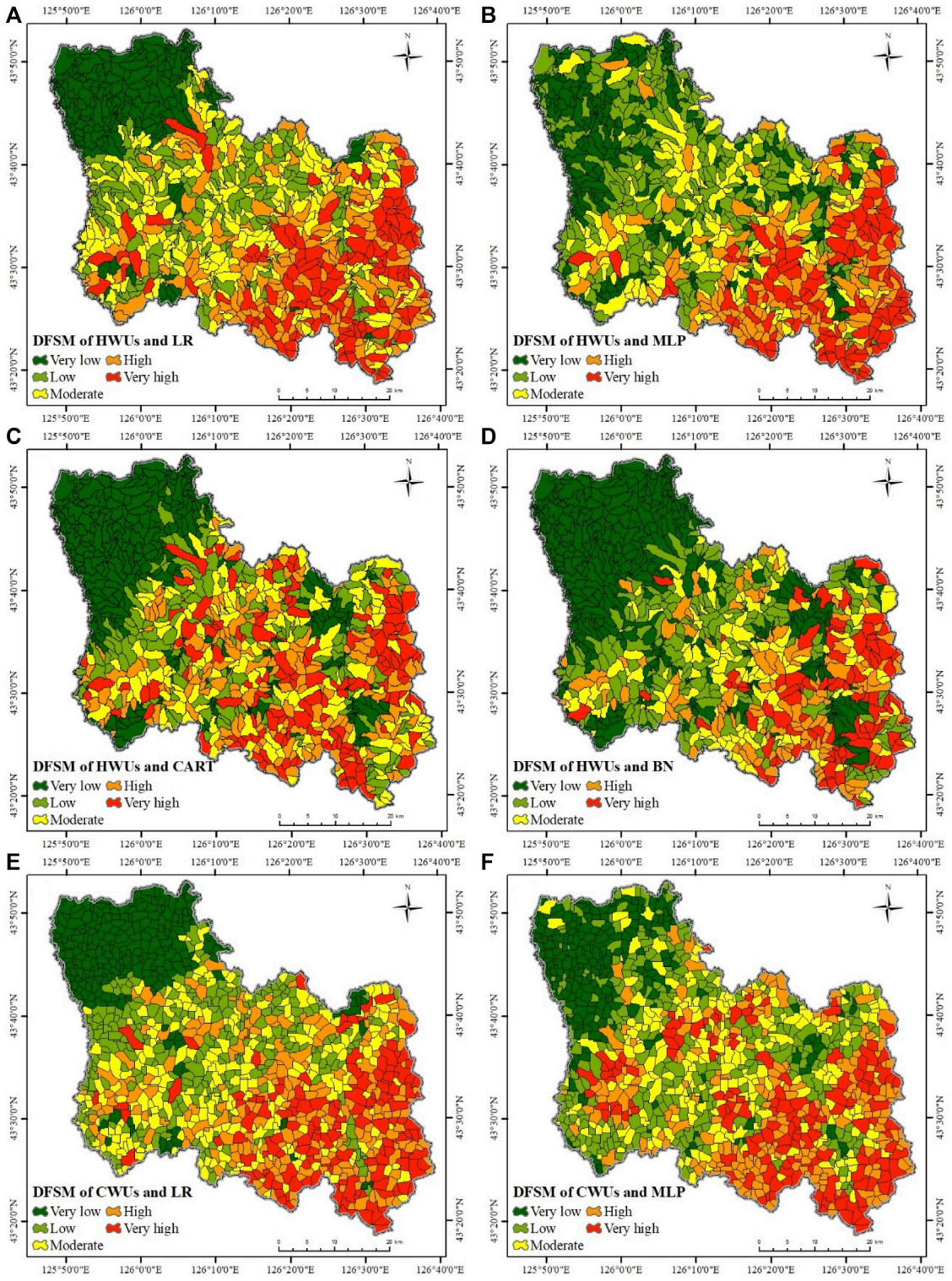


FIGURE 8
(Continued).

$$K = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

$$p_e = \frac{A \times a + B \times b}{n^2} \tag{10}$$

where K is the Kappa coefficient; p_0 is overall classification accuracy, namely, ACC; A is the actual number of debris flows, and it is also the sum of TP and FN (Table 4); B is the actual number of non-debris flows, and it is also the sum of FP

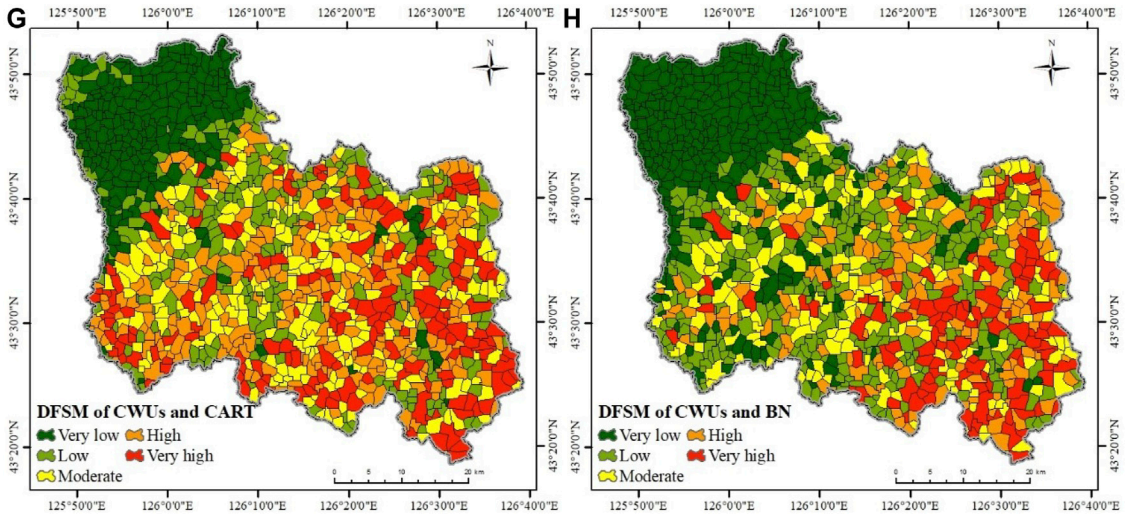


FIGURE 8
(Continued). Eight DFSMs: **(A)** DFSM of HWUs and LR; **(B)** DFSM of HWUs and MLP; **(C)** DFSM of HWUs and CART; **(D)** DFSM of HWUs and BN; **(E)** DFSM of CWUs and LR; **(F)** DFSM of CWUs and MLP; **(G)** DFSM of CWUs and CART and **(H)** DFSM of CWUs and BN.

and TN; a is the predicted number of debris flows, and it is also the sum of TP and FP; b is the predicted number of non-debris flows, and it is also the sum of FN and TN; n is the total number of samples, and it is also the sum of A and B or a and b. The K varies from 0.0 to 1.0, the higher the K value is, the better the classification accuracy of the model.

The receiver operating characteristic (ROC) curve and area under the curve (AUC) can compare the prediction performance of different classifiers (Akgun et al., 2012). The abscissa and ordinate of the ROC are the false-positive rate (FPR) and true-positive rate (TPR) respectively. They can be obtained from the following equations (Pourghasemi et al., 2013):

$$FPR = \frac{FP}{FP + TN} \tag{11}$$

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

AUC represents the quality of models that reliably predict the occurrence or non-occurrence of debris flows. The AUC varies from 0.5 to 1.0, and the higher the AUC value is, the better the prediction performance of the model.

5 Results and validation

5.1 Model parameters

The whole analysis process was implemented in IBM SPSS software (Sun et al., 2019; Sun et al., 2021). For LR, the forward step mode was adopted to screened variables. For BN model, the mechanism type was Tree Augmented naive Bayes (TAN), and Bayesian adjustment of small cell count was selected as a parameter learning method. For MLP, one hidden layer was selected, and the maximum training time was used as the termination rule. For CART, the maximum tree depth was set to 10, and percentage was used as the termination rule. Other parameters are default.

5.2 Contribution of debris flow influencing factors

The Chi-Squared statistic was employed to identify the most important factors affecting the occurrence of debris flows in the study area (Ghasemian et al., 2022a). Figure 7 shows that landforms have the highest impact (128.5) on debris flows in the study area, followed by relative height difference and SPI (114.5), watershed slope (110.5), TWI (82.5), precipitation (74.5), lithology (60), land use (43.5), watershed area (27.5), river density (26.5), mean curvature (24), NDVI (23), and fault density (21.5).

5.3 Spatial datasets for model building

According to field surveys and historical materials, a total of 129 debris flows were collected. Meanwhile, 129 non-debris flows were selected, which were at least 500 m away from the nearest debris flow (Figure 4) (Dou et al., 2019; Sun et al., 2020). Assigned 1 and 0 for debris flows and non-debris flows, respectively. The FRs of the thirteen influencing factors shown in Table 2 were taken as the input variables, and the debris flows and non-debris flows were taken as the output variables. For all 258 samples, 70% (n = 180) were selected randomly for training data, which were used to create the DFSM models. The remaining 30% (n = 78) were used as testing data, which were applied to validate the DFSM models. Based on two types of watershed units (HWUs and CWUs) and four models (LR, MLP, CART and BN), eight DFSMs of Yongji county were completed.

	HM	HL	HC	HB	CM	CL	CC	CB
VH	16.36	19.27	17.81	12.85	18.39	18.03	19.96	14.34
H	17.00	20.10	17.42	13.82	19.46	20.55	21.51	14.75
M	18.96	23.79	19.33	14.87	20.29	21.94	20.08	17.26
L	28.33	17.78	17.70	21.75	25.31	20.24	19.63	26.52
VL	19.34	19.07	27.74	36.70	16.56	19.25	18.81	27.13

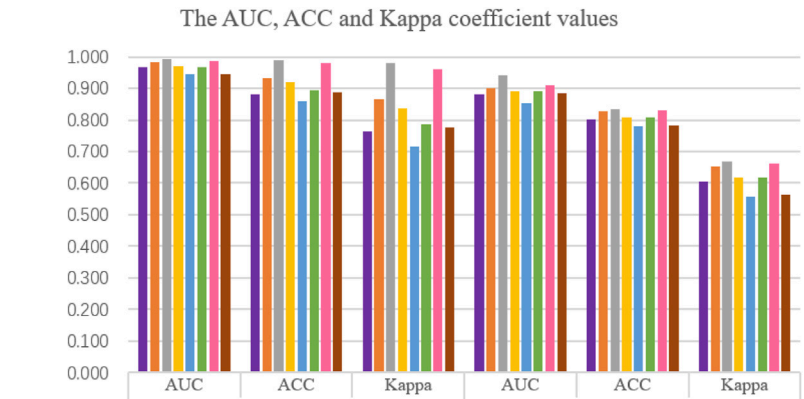
	HM	HL	HC	HB	CM	CL	CC	CB
VH	67.44	76.74	76.74	72.09	67.44	70.54	62.79	64.34
H	14.73	12.40	13.18	18.60	20.16	16.28	22.48	24.81
M	14.73	8.53	6.98	6.98	9.30	11.63	13.95	9.30
L	2.33	2.33	3.10	2.33	3.10	1.55	0.78	1.55
VL	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00

	HM	HL	HC	HB	CM	CL	CC	CB
VH	4.12	3.98	4.31	5.61	3.67	3.91	3.15	4.49
H	0.87	0.62	0.76	1.35	1.04	0.79	1.04	1.68
M	0.78	0.36	0.36	0.47	0.46	0.53	0.69	0.54
L	0.08	0.13	0.18	0.11	0.12	0.08	0.04	0.06
VL	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00

FIGURE 9
The classification of DFSMs and debris flow density: **(A)** area ratio; **(B)** debris flows ratio; **(C)** debris flow density.

5.4 Eight DFSMs

In this paper, IBM SPSS software was chosen to build the debris flow susceptibility predictive models. The model outputs are the debris flow susceptibility indices of all watershed units in the study area. Debris flow susceptibility indices are the probability of debris flow occurrence which varies from 0 to 1 (Xiong et al., 2020). Based on the ArcGIS software, the debris flow susceptibility indices were



	AUC	ACC	Kappa	AUC	ACC	Kappa
	Training			Testing		
HWU_LR	0.966	0.881	0.762	0.882	0.802	0.604
HWU_MLP	0.982	0.932	0.864	0.902	0.826	0.652
HWU_CART	0.991	0.990	0.980	0.942	0.834	0.668
HWU_BN	0.970	0.919	0.838	0.892	0.808	0.617
CWU_LR	0.946	0.858	0.716	0.851	0.779	0.558
CWU_MLP	0.967	0.893	0.787	0.890	0.809	0.618
CWU_CART	0.985	0.980	0.960	0.909	0.832	0.663
CWU_BN	0.946	0.889	0.778	0.884	0.781	0.563

FIGURE 10
The AUC, ACC, and Kappa coefficient values of the eight models for the training and testing data.

converted into raster format to produce the debris flow susceptibility map. Quantile classification was applied to divide the final maps into five classes, namely, very low susceptibility (VL), low susceptibility (L), moderate susceptibility (M), high susceptibility (H), and very high susceptibility (VH). (Martha et al., 2013; Hussin et al., 2016; Steger et al., 2017).

As shown in Figure 8, the susceptibility distributions of the eight models have common characteristics. Very high and high susceptibility areas are mainly distributed in the southeast, moderate susceptibility areas are mainly distributed in the middle, and very low and low susceptibility areas are mainly distributed in northwestern of Yongji county, which is consistent with previous research results (Qin et al., 2019). The landform in the southeast of the study area is mainly middle mountains, and the land use is mainly forest and farmland. The watershed units distributed in the southeast have large relative height differences and slopes, which leads to frequent debris flow disasters. The precipitation decreases from southeast to northwest, which is consistent with the susceptibility distribution. The lithology in southeastern Yongji county is hard massive rock, mainly granite. Weathered granite is a component of debris flows, which increases the density and destructive power of debris flows (Figures 2B–E).

For the eight DFSMs, the area ratios of the five susceptibility classes (very high, high, moderate, low, and very low) were 12.85–19.96, 13.82–21.51, 14.87–23.79, 17.70–28.33, and 16.56%–36.70%, respectively (Figure 9A); The debris flow ratios of the five susceptibility classes were 62.79–76.74, 12.40–24.81, 6.98–14.73,0.78%–3.10% and 0%–0.78%, respectively (Figure 9B). As shown in Figure 9C, the debris flow density was calculated to evaluate the performance of the DFSMs, that is, the ratio of debris flow percentage to area percentage on each susceptible class (Pham et al., 2016). The maximum values of the debris flow density of the eight models appear in the very high susceptibility class, varying from 3.15 to 5.61. The minimum values all appear in the very low susceptibility class, varying from 0.00 to 0.04. The debris flow density increases gradually from a very low class to a very high class, which provides a good visualization of the spatial predictions of debris flows (Pham et al., 2017; Asadi et al., 2022).

5.5 Validation and comparison of the models

Model validation is a vital step in disaster susceptibility mapping (Wang et al., 2022). By considering the three commonly used performance metrics of ACC, AUC and Kappa, eight models were verified. The AUC, ACC and Kappa coefficient values of the eight models on the training and testing data are shown in Figure 10.

In the training phase, when HWUs were used as the mapping unit, the ACC stated that HWUs_CART model had the highest value (0.990), followed by HWUs_MLP (0.932), HWUs_BN (0.919) and HWUs_LR (0.881). It showed that the HWUs_CART model can correctly classify the debris flow and non-debris flow locations as debris flow and non-debris flow situations respectively. The highest and lowest Kappa values were 0.980 and 0.762, respectively for the HWUs_CART and HWUs_LR. Meanwhile, HWUs_MLP (0.864) and HWUs_BN (0.838) was ranked in other positions. In terms of AUC, results indicated that the HWUs_CART model with a value of 0.991 had higher performance than the HWUs_MLP (0.982), HWUs_BN (0.970) and HWUs_LR (0.966). When CWUs was used as the mapping unit, the ACC, Kappa and AUC values of the CWUs_CART model were 0.980, 0.960 and 0.985, which showed that the performance of the CWUs_CART model was the highest,

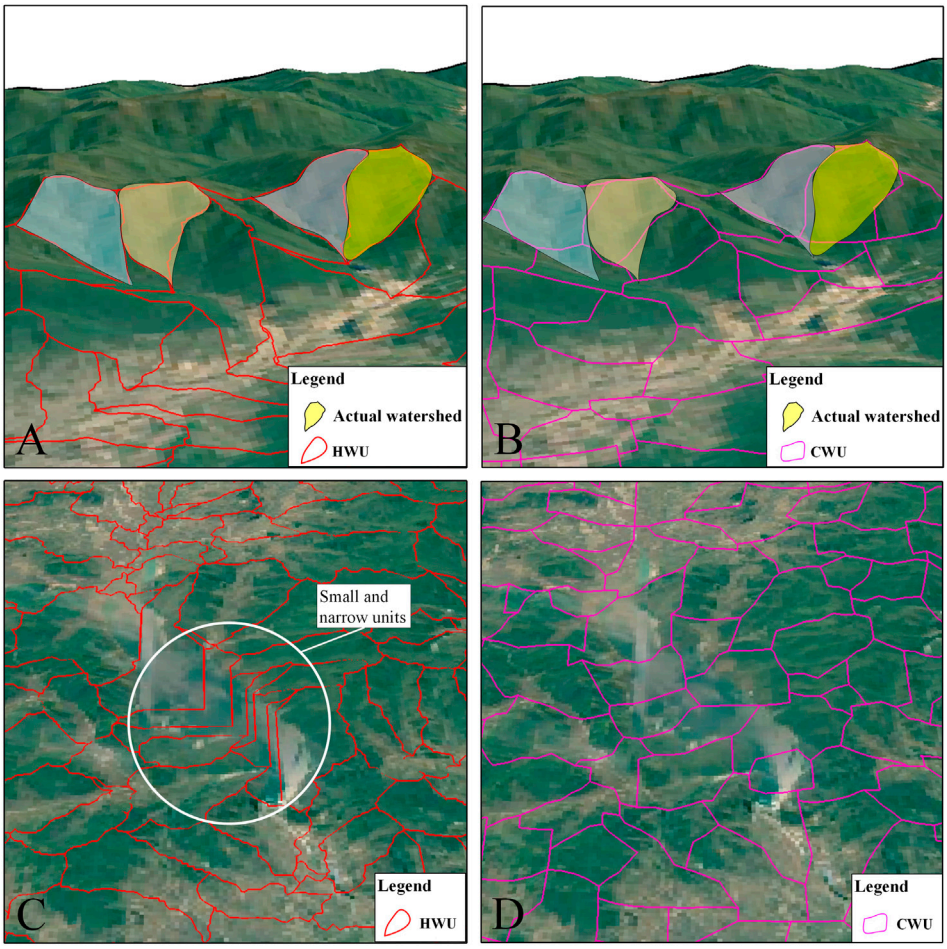


FIGURE 11
Watershed unit classification results comparison: **(A)** The boundaries of HWUs are relatively consistent with reality in mountainous areas; **(B)** The boundaries of CWUs do not match well with the actual situation in mountainous areas; **(C)** Small and narrow units in flat areas of HWUs; and **(D)** The division of CWUs in flat areas is relatively satisfactory.

TABLE 5 AUC values of testing data in different studies.

Studies	Method	Mapping units	AUC
This study	LR	HWUs	0.882
	MLP	HWUs	0.902
	CART	HWUs	0.942
	BN	HWUs	0.892
Qin et al. (2019)	FR	HWUs	0.879
Qiao et al. (2021)	AHP	HWUs	0.812

followed by the CWUs_MLP (0.893, 0.787, 0.967), CWUs_BN (0.899, 0.778, and 0.946) and CWUs_LR (0.858, 0.716, and 0.946) (Figure 10). Although the results showed the excellent performance for all the four algorithms, the CART had the highest ability in debris flow classification and susceptibility mapping in the study area. In terms of watershed unit, ACC, Kappa and AUC values decreased when HWUs was replaced by CWUs, indicating that HWUs were more suitable for DFSM in the study area than CWUs.

Right side of Figure 10 showed the prediction capabilities of the eight models based on testing dataset. These results are very important for evaluating the applicability and robustness of the models. When HWUs were used as the mapping unit, the highest value of ACC was 0.834 for the HWUs_CART model, next for the HWUs_MLP (0.826), HWUs_BN (0.808) and HWUs_LR (0.802) models. The Kappa for the HWUs_CART model was 0.668 as the highest value, whereas this value was 0.652, 0.617, and 0.604 for HWUs_MLP, HWUs_BN, and HWUs_LR, respectively. The highest and lowest AUC values were 0.942 and 0.882, respectively for the HWUs_CART and HWUs_LR. Meanwhile, HWUs_MLP (0.902) and HWUs_BN (0.892) was ranked in other positions. Correspondingly, ACC, Kappa and AUC from CWUs were shown in Figure 9, which indicated a similar result with HWUs. CART model resulted in the highest ACC, Kappa and AUC values of 0.832, 0.663, and 0.909, which manifested it is the best model for the study area. At the same time, the HWU-based models had better performance than the CWU models for DFSM in the study area.

5.6 One-way ANOVA test

The results of the models are tested by one-way ANOVA in SPSS. For HWUs, there are significant differences between CART and each of the three methods (LR, MLP, and BN). There are no significant differences among LR, MLP, and BN. For CWUs, there are no significant differences between MLP and each of the two methods (LR and BN). There are significant differences between the other methods.

6 Discussion

6.1 Watershed unit classification processes and results comparison

As shown in Figure 3, the extraction processes of HWUs are more complex than those of CWUs, because HWUs require six steps while CWUs require five steps. Model builder in ArcGIS is a workflow that connects a series of geoprocessing tools (Qin et al., 2019). It takes the output of one tool as the input of the other tool. Model builder can greatly reduce operation time and improve work efficiency. We had built two workflows for the processes of extracting HWUs and CWUs in the model builder. Experiments on two types of watershed units showed that HWUs extraction required 17 s, while CWUs extraction required only 3 s. In addition, for the division of HWUs, the influence of DEM resolution and flow threshold needs to be considered, while for CWUs, only DEM resolution needs to be considered. In summary, it takes more time and effort to extract HWUs than CWUs.

There are also significant differences between the two types of watershed units in the classification results. As shown in Figure 4, HWUs are mostly strip-shaped and widely different in size, while CWUs are nearly square and relatively uniform in size. The watershed unit boundaries extracted by the hydrological analysis method in areas with obvious topographic changes are relatively consistent with reality (Figure 11A). However, there are many small and narrow units in flat areas, because there is no clear flow direction in flat areas for hydrological analysis (Massimiliano et al., 2016) (Figure 11C). For the curvature method, the division of watershed units in flat areas is relatively satisfactory, and there is no parallel line problem similar to the hydrological analysis method (Figure 11D). However, in areas with obvious topographic changes, the boundaries of watershed units do not match well with the actual situation (Figure 11B).

6.2 Comparison of DFSMs based on different evaluation models

Four models, LR, MLP, CART and BN, were used to complete the DFSMs of Yongji county in this study. Figure 10 shows the predictive ability of the eight models. When HWUs were used as mapping units, CART had the highest evaluation criteria with AUC, ACC and Kappa values of 0.991, 0.990, and 0.980 respectively, followed by MLP (0.982, 0.932, 0.864), BN (0.970, 0.919, 0.838) and LR (0.966, 0.881, 0.762) in the training stage. For the testing stage, the CART had the highest prediction accuracy with AUC, ACC and Kappa values of 0.942, 0.834 and 0.668 respectively, followed by MLP (0.902, 0.826, 0.652), BN (0.892, 0.808, 0.617) and LR (0.882, 0.802, 0.604). When CWUs were used as mapping units, the evaluation results showed the same trend as HWUs. The comparisons of the four evaluation models show that the CART had the best predictive ability over the other three models. The current research was in agreement with previous research results. Wang et al. (2015) analyzed landslide susceptibility based on five mathematical models (artificial neural network, frequency ratio, CART, LR and weights of evidence methods) and three sampling strategies. They indicated the results obtained from CART show steady prediction power with an AUC value larger than 0.7. Felicísimo et al. (2012) indicated that the CART is one of the most predictive models with the AUC value of 0.77. Using random forest (RF), boosted regression tree (BRT), classification and regression tree (CART), and general linear (GLM), Youssef et al. (2015) found the success rate for CART was 0.816 and for the prediction rate the CART was the highest with a value of 0.862. CART represents information in an intuitive and easy visual way, and is widely used in many fields (Bevilacqua et al., 2003; Malinowska 2014; Kim et al., 2015; Youssef et al., 2015; Yang et al., 2016).

Several studies have been conducted in similar areas. Selecting the frequency ratio (FR) model as the statistical method, Qin et al. (2019) explored the accuracy and practicability of HWUs and grid cell units (GCUs) in evaluating debris flow susceptibility in Yongji county. Qiao et al. (2021) built debris flow susceptibility models *via* the analytical hierarchy process (AHP) method and generated maps of Yongji county. The AUC values of the testing data in different studies are shown in Table 5. When the HWUs were chosen as mapping units, the AUC values of the DFSMs based on AHP, FR, LR, BN, MLP and CART were 0.812, 0.879, 0.882, 0.892, 0.902, and 0.942 respectively. The main difference among these DFSMs is the selection of different evaluation models, which indicates that machine learning models can improve the prediction accuracy of DFSMs. These results are consistent with previous studies, indicating that machine learning models are more suitable for DFSM than heuristic and general statistical models (Huang et al., 2020; Sun et al., 2021).

6.3 Comparison of DFSMs based on different watershed units

The selection of mapping units is one of the key issues for the rationality and correctness of disaster susceptibility mapping (Van Den Eeckhaut et al., 2009; Chen et al., 2019; Sun et al., 2020). The impact of different mapping units on disaster susceptibility mapping is greater than that of statistical methods (Zezere et al., 2017). Although many studies have tried to compare different evaluation models for disaster susceptibility mapping (Achour et al., 2018; Liang et al., 2020; Xiong et al., 2020; Dash et al., 2022; Qiu et al., 2022), very few studies have considered different mapping units. Qin et al. (Qin et al., 2019) explored the effect of grid cell unit and HWUs on the susceptibility mapping of debris flow, they found HWUs can reflect the geological and geomorphic environmental conditions of a debris flow accurately and perfectly. Li et al. (Li et al., 2021) discussed the influence of four different HWUs on debris flow susceptibility assessment results.

The results show that the appropriate watershed division scheme can obtain more reasonable results. In this study, HWUs and CWUs were selected to map debris flow susceptibility. When the CART was selected as the machine learning model, the HWUs generated high AUC, ACC, and Kappa for training data (0.991, 0.990 and 0.980) compared to the CWUs (0.985, 0.980, and 0.960). For testing data, the AUC, ACC, and Kappa of HWUs were 0.942, 0.834, and 0.668, respectively. The AUC, ACC, and Kappa of CWUs were 0.909, 0.832, and 0.663, respectively. The results suggest that the HWU model has a higher debris flow prediction performance than the CWU model. The same trend can be observed in the LR, MLP, and BN models. Therefore, the HWU-based model is superior to the CWU-based model in debris flow susceptibility assessment due to higher training and testing accuracy.

As described in "6.1 Watershed unit classification processes and results comparison," compared with CWUs, HWUs agree well with the actual watershed units in mountainous areas, but small and narrow units appear in plain areas. Since the frequency of debris flows in mountainous areas is much higher than that in plain areas, the division of watershed units in mountainous areas is more important than that in plain areas. Therefore, the HWU model is more practical than the CWU model. CWUs can also represent the distribution of watersheds and can be used as an alternative scheme.

Although this paper discussed the application of two types of watershed units in DFSM and obtained positive results, there are some limitations: 1) the number of debris flows is small, and 2) only HWUs with a threshold of 1,000 and CWUs with a resolution of 300 are selected for comparison. In future research, we will constantly update the debris flow database to improve the data quality. Moreover, it is necessary to explore the similarities and differences of multiscale watershed units in DFSM.

7 Conclusion

This paper mainly explored the influence of using different watershed units (HWUs and CWUs) in debris flow susceptibility assessment models. LR, MLP, CART, and BN were chosen as evaluation models to avoid the model uncertainty caused by different models. Yongji county, with 129 recorded debris flows and 13 related influencing factors, was used as the study area and eight DFSMs were produced.

The DFSM results showed that CART has the best predictive ability over the other three models through the analysis of AUC, ACC and Kappa. By using Model Builder in ArcGIS, 1,092 HWUs and 1,211 CWUs were extracted. Compared with HWUs, the extraction process of CWUs is simpler. For the results of watershed unit division, HWUs have more advantages in areas with undulating terrain, but they are not satisfactory in areas with flat terrain. CWUs perform well in flat areas but do not match the actual watershed boundaries in areas with undulating terrain. Since debris flows mostly occur in mountainous areas, the DFSM based on HWUs is more accurate and practical than that based on CWUs. In addition, the AUC, ACC and Kappa showed that the HWU-based model has remarkably higher debris flow prediction performance than CWUs. This result means that the HWUs are more effective in debris flow susceptibility assessment of the study area. The CWU-based model can also reflect the spatial distribution probability of debris flows in the study area overall and can be used as an alternative model. Further studies should propose a more appropriate watershed unit for DFSM.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JL: Conceptualization, methodology, formal analysis, investigation, writing, original draft, writing-review and editing. SUQ: Methodology, validation, resources, data curation, writing -original draft, visualization, project administration, funding acquisition. JC: Validation, investigation, supervision, project administration. SGQ: Investigation, supervision. JGY: Software, data curation. XZ: Conceptualization, supervision. RC: Software, supervision. JHY: Investigation, data curation.

Funding

This work was funded by the National Natural Science Foundation of China under Grants 41977221 and 41972267, and in part by the Jilin Provincial Science and Technology Department (Grant No. 20190303103SF).

Acknowledgments

The authors are also thankful to the reviewers for their valuable feedback on the manuscript.

Conflict of interest

JL, XZ, RC, and JHY were employed by China Water Resources Bei Fang Investigation, Design & Research Co. LTD.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Achour, Y., Garçia, S., and Cavaleiro, V. (2018). GIS-based spatial prediction of debris flows using logistic regression and frequency ratio models for Zêzere River basin and its surrounding area, Northwest Covilhã, Portugal. *Arabian J. Geosciences* 11, 550. doi:10.1007/s12517-018-3920-9

Aditian, A., Kubota, T., and Shinohara, Y. (2018). Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* 318, 101–111. doi:10.1016/j.geomorph.2018.06.006

Akgun, A., Kincal, C., and Pradhan, B. (2012). Application of remote sensing data and GIS for landslide risk assessment as an environmental threat to Izmir city (west Turkey). *Environ. Monit. Assess.* 184, 5453–5470. doi:10.1007/s10661-011-2352-8

Althuwaynee, O. F., Pradhan, B., Park, H. J., and Lee, J. H. (2014). A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides* 11, 1063–1078. doi:10.1007/s10346-014-0466-0

Arabameri, A., Saha, S., Roy, J., Chen, W., Blaschke, T., and Tien Bui, D. (2020). Landslide susceptibility evaluation and management using different machine learning methods in the gallicash river watershed, Iran. *Remote Sens.* 12, 475. doi:10.3390/rs12030475

Asadi, M., Goli Mokhtari, L., Shirzadi, A., Shahabi, H., and Bahrami, S. (2022). A comparison study on the quantitative statistical methods for spatial prediction of shallow landslides (case study: Yozidar-Degaga Route in Kurdistan Province, Iran). *Environ. Earth Sci.* 81, 51. doi:10.1007/s12665-021-10152-4

Ayalew, L., and Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Balamurugan, G., Ramesh, V., and Touthang, M. (2016). Landslide susceptibility zonation mapping using frequency ratio and fuzzy gamma operator models in part of NH-39, Manipur, India. *Nat. Hazards* 84, 465–488. doi:10.1007/s11069-016-2434-6

Bălteanu, D., Chendeş, V., Sima, M., and Enciu, P. (2010). A country-wide spatial assessment of landslide susceptibility in Romania. *Geomorphology* 124, 102–112. doi:10.1016/j.geomorph.2010.03.005

Bevilacqua, M., Braglia, M., and Montanari, R. (2003). The classification and regression tree approach to pump failure rate analysis. *Reliab. Eng. Syst. Saf.* 79, 59–67. doi:10.1016/s0951-8320(02)00180-1

Bregoli, F., Medina, V., Chevalier, G., Hürlimann, M., and Bateman, A. (2014). Debris-flow susceptibility assessment at regional scale: Validation on an alpine environment. *Landslides* 12, 437–454. doi:10.1007/s10346-014-0493-x

Breiman, L. F. F., Olshen, R., and Stone, C. (1984). Classification and regression trees. Wadsworth. *Biometrics* 40, 358.

Cama, M., Conoscenti, C., Lombardo, L., and Rotigliano, E. (2016). Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: A test in the giampilieri catchment (sicily, Italy). *Environ. Earth Sci.* 75, 238. doi:10.1007/s12665-015-5047-6

Cao, J., Zhang, Z., Wang, C., Liu, J., and Zhang, L. (2019). Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *Catena* 175, 63–76. doi:10.1016/j.catena.2018.12.013

Carranza, E. J. M. (2014). Data-Driven evidential belief modeling of mineral potential using few prospects and evidence with missing values. *Nat. Resour. Res.* 24, 291–304. doi:10.1007/s11053-014-9250-z

Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W., et al. (2020). Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models. *Remote Sens.* 12, 502. doi:10.3390/rs12030502

Chen, J., Dai, Z., Dong, S., Zhang, X., Sun, G., Wu, J., et al. (2022). Integration of deep learning and information theory for designing monitoring networks in heterogeneous aquifer systems. *Water Resour. Res.* 58, 429. doi:10.1029/2022wr032429

Chen, J., Dai, Z., Yang, Z., Pan, Y., Zhang, X., Wu, J., et al. (2021). An improved tandem neural network architecture for inverse modeling of multicomponent reactive transport in porous media. *Water Resour. Res.* 57, 595. doi:10.1029/2021wr030595

Chen, W., Shahabi, H., Shirzadi, A., Hong, H., Akgun, A., Tian, Y., et al. (2018). Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* 78, 4397–4419. doi:10.1007/s10064-018-1401-8

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., et al. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151, 147–160. doi:10.1016/j.catena.2016.11.032

Chen, Z., Liang, S., Ke, Y., Yang, Z., and Zhao, H. (2019). Landslide susceptibility assessment using different slope units based on the evidential belief function model. *Geocarto Int.* 35, 1641–1664. doi:10.1080/10106049.2019.1582716

Colkesen, I., Sahin, E. K., and Kavzoglu, T. (2016). Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* 118, 53–64. doi:10.1016/j.jafrearsci.2016.02.019

Constantin, M., Bednarik, M., Jurchescu, M. C., and Vlaicu, M. (2010). Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). *Environ. Earth Sci.* 63, 397–406. doi:10.1007/s12665-010-0724-y

Dash, R. K., Falae, P. O., and Kanungo, D. P. (2022). Debris flow susceptibility zonation using statistical models in parts of northwest Indian himalayas—Implementation, validation, and comparative evaluation. *Nat. Hazards* 111, 2011–2058. doi:10.1007/s11069-021-05128-3

Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., et al. (2012). Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* 65, 135–165. doi:10.1007/s11069-012-0347-6

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi:10.1111/j.1600-0587.2012.07348.x

Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., et al. (2019a). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* 662, 332–346. doi:10.1016/j.scitotenv.2019.01.221

Dou, Q., Qin, S., Zhang, Y., Ma, Z., Chen, J., Qiao, S., et al. (2019b). A method for improving controlling factors based on information fusion for debris flow susceptibility mapping: A case study in Jilin Province, China. *Entropy (Basel)* 21, 695. doi:10.3390/e21070695

Dragut, L., and Eisank, C. (2011). Object representations at multiple scales from digital elevation models. *Geomorphol. (Amst)* 129, 183–189. doi:10.1016/j.geomorph.2011.03.003

Du, G., Zhang, Y., Yang, Z., Guo, C., Yao, X., and Sun, D. (2018). Landslide susceptibility mapping in the region of eastern himalayan syntaxis, Tibetan plateau, China: A comparison between analytical hierarchy process information value and logistic regression-information value methods. *Bull. Eng. Geol. Environ.* 78, 4201–4215. doi:10.1007/s10064-018-1393-4

Esper Angillieri, M. Y. (2020). Debris flow susceptibility mapping using frequency ratio and seed cells, in a portion of a mountain international route, Dry Central Andes of Argentina. *Catena* 189, 104504. doi:10.1016/j.catena.2020.104504

Fausto Guzzetti, A. C., Cardinali, M., Reichenbach, P., and Reichenbach, P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, central Italy. *Geomorphology* 31, 181–216. doi:10.1016/S0169-555X(99)00078-1

Felicísimo, Á. M., Cuartero, A., Remondo, J., and Quirós, E. (2012). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslides* 10, 175–189. doi:10.1007/s10346-012-0320-1

Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Geertsema, M., et al. (2022a). Application of a novel hybrid machine learning algorithm in shallow landslide susceptibility mapping in a mountainous area. *Front. Environ. Sci.* 10, 897254. doi:10.3389/fenvs.2022.897254

Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Kress, V. R., et al. (2022b). A robust deep-learning model for landslide susceptibility mapping: A case study of kurdistan Province, Iran. *Sensors (Basel)* 22, 1573. doi:10.3390/s22041573

Hadmoko, D. S., Lavigne, F., and Samodra, G. (2017). Application of a semiquantitative and GIS-based statistical model to landslide susceptibility zonation in Kayangan Catchment, Java, Indonesia. *Nat. Hazards* 87, 437–468. doi:10.1007/s11069-017-2772-z

Han, L., Zhang, J., Zhang, Y., and Lang, Q. (2019). Applying a series and parallel model and a bayesian networks model to produce disaster chain susceptibility maps in the changbai mountain area, China. *Water* 11, 2144. doi:10.3390/w11102144

He, S., Pan, P., Dai, L., Wang, H., and Liu, J. (2012). Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology* 171-172, 30–41. doi:10.1016/j.geomorph.2012.04.024

Hitoshi Saito, D. N., and Matsuyama, H. (2009). Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology* 109, 108–121. doi:10.1016/j.geomorph.2009.02.026

Hong, H., Pourghasemi, H. R., and Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in lianhua county (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 259, 105–118. doi:10.1016/j.geomorph.2016.02.012

Hong, H., Pradhan, B., Xu, C., and Tien Bui, D. (2015). Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression,

alternating decision tree and support vector machines. *Catena* 133, 266–281. doi:10.1016/j.catena.2015.05.019

Hong, H., Tsangaratos, P., Ilia, I., Loupasakis, C., and Wang, Y. (2020). Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. *Sci. Total Environ.* 742, 140549. doi:10.1016/j.scitotenv.2020.140549

Horton, P., Jaboyedoff, M., Rudaz, B., and Zimmermann, M. (2013). Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. *Nat. Hazards Earth Syst. Sci.* 13, 869–885. doi:10.5194/nhess-13-869-2013

Hu, W., Xu, Q., Rui, C., Huang, R. Q., van Asch, T. W. J., Zhu, X., et al. (2014). An instrumented flume to investigate the initiation mechanism of the post-earthquake huge debris flow in the southwest of China. *Bull. Eng. Geol. Environ.* 74, 393–404. doi:10.1007/s10064-014-0627-3

Huang, F., Cao, Z., Guo, J., Jiang, S. H., Li, S., and Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena* 191, 104580. doi:10.1016/j.catena.2020.104580

Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., et al. (2016). Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* 253, 508–523. doi:10.1016/j.geomorph.2015.10.030

Ilia, I., and Tsangaratos, P. (2015). Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. *Landslides* 13, 379–397. doi:10.1007/s10346-015-0576-3

Jaafari, A., Panahi, M., Pham, B. T., Shahabi, H., Bui, D. T., Rezaie, F., et al. (2019). Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* 175, 430–445. doi:10.1016/j.catena.2018.12.033

Jiang, H., Zou, Q., Zhou, B., Hu, Z., Li, C., Yao, S., et al. (2022). Susceptibility assessment of debris flows coupled with ecohydrological activation in the eastern qinghai-tibet plateau. *Remote Sens.* 14, 1444. doi:10.3390/rs14061444

Kavzoglu, T., and Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *Int. J. Remote Sens.* 24, 4907–4938. doi:10.1080/0143116031000114851

Kim, K. N., Kim, D. W., and Jeong, M. A. (2015). The usefulness of a classification and regression tree algorithm for detecting perioperative transfusion-related pulmonary complications. *Transfusion* 55, 2582–2589. doi:10.1111/trf.13202

Lee, S. (2007). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* 26, 1477–1491. doi:10.1080/01431160412331331012

Lee, S., and Pradhan, B. (2006). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4, 33–41. doi:10.1007/s10346-006-0047-y

Lei, T. C., Wan, S., Chou, T. Y., and Pai, H. C. (2010). The knowledge expression on debris flow potential analysis through PCA + LDA and rough sets theory: A case study of chen-yu-lan watershed, nantou, taiwan. *Environ. Earth Sci.* 63, 981–997. doi:10.1007/s12665-010-0775-0

Li, D., Huang, F., Yan, L., Cao, Z., Chen, J., and Ye, Z. (2019). Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: Comparisons with multilayer-perceptron-only, BP neural network, and information value models. *Appl. Sci.* 9, 3664. doi:10.3390/app9183664

Li, Y., Wang, H., Chen, J., and Shang, Y. (2017). Debris flow susceptibility assessment in the Wudongde Dam area, China based on rock engineering system and fuzzy C-means algorithm. *Water* 9, 669. doi:10.3390/w9090669

Li, Z., Chen, J., Tan, C., Zhou, X., Li, Y., and Han, M. (2021). Debris flow susceptibility assessment based on topo-hydrological factors at different unit scales: A case study of mentougou district, beijing. *Environ. Earth Sci.* 80, 365. doi:10.1007/s12665-021-09665-9

Liang, Z., Wang, C. M., Zhang, Z. M., and Khan, K. U. J. (2020). A comparison of statistical and machine learning methods for debris flow susceptibility mapping. *Stoch. Environ. Res. Risk Assess.* 34, 1887–1907. doi:10.1007/s00477-020-01851-8

Malinowska, A. (2014). Classification and regression tree theory application for assessment of building damage caused by surface deformation. *Nat. Hazards* 73, 317–334. doi:10.1007/s11069-014-1070-2

Martha, T. R., van Westen, C. J., Kerle, N., Jetten, V., and Vinod Kumar, K. (2013). Landslide hazard and risk assessment using semi-automatically created landslide inventories. *Geomorphology* 184, 139–150. doi:10.1016/j.geomorph.2012.12.001

Massimiliano, A., Ivan, M., Paola, R., Mauro, R., Francesca, A., Federica, F., et al. (2016). Automatic delineation of geomorphological slope units with <tt>r.slopeunits v1.0</tt> and their optimization for landslide susceptibility modeling. *Geosci. Model. Dev. Discuss.* 9, 3975–3991. doi:10.5194/gmd-9-3975-2016

Pham, B. T., Tien Bui, D., Prakash, I., and Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* 149, 52–63. doi:10.1016/j.catena.2016.09.007

Pham, B. T., Tien Bui, D., Prakash, I., and Dholakia, M. B. (2016). Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Nat. Hazards* 83, 97–127. doi:10.1007/s11069-016-2304-2

Pourghasemi, H. R., Moradi, H. R., and Fatemi Aghda, S. M. (2013). Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* 69, 749–779. doi:10.1007/s11069-013-0728-5

Qiao, S. S., Qin, S. W., Sun, J. B., Che, W. C., Yao, J. Y., Su, G., et al. (2021). Development of a region-partitioning method for debris flow susceptibility mapping. *J. Mt. Sci.* 18, 1177–1191. doi:10.1007/s11629-020-6497-1

Qin, S., Lv, J., Cao, C., Ma, Z., Hu, X., Liu, F., et al. (2019). Mapping debris flow susceptibility based on watershed unit and grid cell unit: A comparison study. *Geomatics, Nat. Hazards Risk* 10, 1648–1666. doi:10.1080/19475705.2019.1604572

Qiu, C., Su, L., Zou, Q., and 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. doi:10.1016/j.scitotenv.2021.151752

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Rev.* 180, 60–91. doi:10.1016/j.earscirev.2018.03.001

Romstad, B., and Etzelmüller, B. (2012). Mean-curvature watersheds: A simple method for segmentation of a digital elevation model into terrain units. *Geomorphology* 139-140, 293–302. doi:10.1016/j.geomorph.2011.10.031

Rozos, D., Bathrellos, G. D., and Skillodimou, H. D. (2010). Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon landslide susceptibility mapping, using GIS: A case study from the eastern achaia county of peloponnesus, Greece. *Environ. Earth Sci.* 63, 49–63. doi:10.1007/s12665-010-0687-z

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., and Malet, J. P. (2018). Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology* 301, 10–20. doi:10.1016/j.geomorph.2017.10.018

Shi, M., Chen, J., Song, Y., Zhang, W., Song, S., and Zhang, X. (2015). Assessing debris flow susceptibility in Heshigten Banner, Inner Mongolia, China, using principal component analysis and an improved fuzzy C-means algorithm. *Bull. Eng. Geol. Environ.* 75, 909–922. doi:10.1007/s10064-015-0784-z

Shirani, K., Pasandi, M., and Arabameri, A. (2018). Landslide susceptibility assessment by dempster–shafer and index of entropy models, sarkhoun basin, southwestern Iran. *Nat. Hazards* 93, 1379–1418. doi:10.1007/s11069-018-3356-2

Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., et al. (2012). Susceptibility assessment of earthquake-induced landslides using bayesian network: A case study in beichuan, China. *Comput. Geosciences* 42, 189–199. doi:10.1016/j.cageo.2011.09.011

Steger, S., Brenning, A., Bell, R., and Glade, T. (2017). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* 14, 1767–1781. doi:10.1007/s10346-017-0820-0

Sun, J., Qin, S., Qiao, S., Chen, Y., Su, G., Cheng, Q., et al. (2021). Exploring the impact of introducing a physical model into statistical methods on the evaluation of regional scale debris flow susceptibility. *Nat. Hazards* 106, 881–912. doi:10.1007/s11069-020-04498-4

Sun, X., Chen, J., Han, X., Bao, Y., Zhan, J., and Peng, W. (2019). 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. doi:10.1007/s10064-019-01572-5

Sun, X., Chen, J., Han, X., Bao, Y., Zhou, X., and Peng, W. (2020). Landslide susceptibility mapping along the upper jinsha river, south-Western China: A comparison of hydrological and curvature watershed methods for slope unit classification. *Bull. Eng. Geol. Environ.* 79, 4657–4670. doi:10.1007/s10064-020-01849-0

Tien Bui, D., Tuan, T. A., Klempe, H., Pradhan, B., and Revhaug, I. (2015). 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, 361–378. doi:10.1007/s10346-015-0557-6

Vakhshoori, V., Pourghasemi, H. R., Zare, M., and Blaschke, T. (2019). Landslide susceptibility mapping using GIS-based data mining algorithms. *Water* 11, 2292. doi:10.3390/w11112292

Van Den Eeckhaut, M, R. P., Guzzetti, F., Rossi, M., and Poesen, J. (2009). Combined landslide inventory and susceptibility assessment based on different mapping units: An example from the flemish ardennes, Belgium. *Nat. Hazards Earth Syst. Sci.* 9, 507–521. doi:10.5194/nhess-9-507-2009

Wang, F., Xu, P., Wang, C., Wang, N., and Jiang, N. (2017). Application of a GIS-based slope unit method for landslide susceptibility mapping along the longzi river,

southeastern Tibetan plateau, China. *ISPRS Int. J. Geo-Information* 6, 172. doi:10.3390/ijgi6060172

Wang, L. J., Guo, M., Sawada, K., Lin, J., and Zhang, J. (2015). A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosciences J.* 20, 117–136. doi:10.1007/s12303-015-0026-1

Wang, X., Huang, F., Fan, X., Shahabi, H., Shirzadi, A., Bian, H., et al. (2022). Landslide susceptibility modeling based on remote sensing data and data mining techniques. *Environ. Earth Sci.* 81, 50. doi:10.1007/s12665-022-10195-1

Xiong, K., Adhikari, B. R., Stamatopoulos, C. A., Zhan, Y., Wu, S., Dong, Z., et al. (2020). Comparison of different machine learning methods for debris flow susceptibility mapping: A case study in the sichuan Province, China. *Remote Sens.* 12, 295. doi:10.3390/rs12020295

Xu, W., Yu, W., Jing, S., Zhang, G., and Huang, J. (2012). Debris flow susceptibility assessment by GIS and information value model in a large-scale region, Sichuan Province (China). *Nat. Hazards* 65, 1379–1392. doi:10.1007/s11069-012-0414-z

Yalcin, A., Reis, S., Aydinoglu, A. C., and Yomralioglu, T. (2011). A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena* 85, 274–287. doi:10.1016/j.catena.2011.01.014

Yang, T., Gao, X., Sorooshian, S., and Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resour. Res.* 52, 1626–1651. doi:10.1002/2015wr017394

Yao, J., Qin, S., Qiao, S., Liu, X., Zhang, L., and Chen, J. (2022). Application of a two-step sampling strategy based on deep neural network for landslide susceptibility mapping. *Bull. Eng. Geol. Environ.* 81, 148. doi:10.1007/s10064-022-02615-0

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., and Al-Katheeri, M. M. (2015). 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. doi:10.1007/s10346-015-0614-1

Zezere, J. L., Pereira, S., Melo, R., Oliveira, S. C., and Garcia, R. A. C. (2017). Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* 589, 250–267. doi:10.1016/j.scitotenv.2017.02.188

Zhang, W., Chen, J. P., Wang, Q., An, Y., Qian, X., Xiang, L., et al. (2012). Susceptibility analysis of large-scale debris flows based on combination weighting and extension methods. *Nat. Hazards* 66, 1073–1100. doi:10.1007/s11069-012-0539-0

Zou, Q., Cui, P., He, J., Lei, Y., and Li, S. (2019). Regional risk assessment of debris flows in China—an HRU-based approach. *Geomorphology* 340, 84–102. doi:10.1016/j.geomorph.2019.04.027

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 01 June 2023
DOI 10.3389/feart.2023.1157881

OPEN ACCESS

EDITED BY
Zefa Yang,
Central South University, China

REVIEWED BY
Zizheng Guo,
Hebei University of Technology, China
Neelima Satyam,
Indian Institute of Technology Indore, India

*CORRESPONDENCE
Laura Ortiz-Giraldo,
lortiz736@soyudemedellin.edu.co

RECEIVED 03 February 2023
ACCEPTED 24 May 2023
PUBLISHED 01 June 2023

CITATION
Ortiz-Giraldo L, Botero BA and Vega J (2023), An integral assessment of landslide dams generated by the occurrence of rainfall-induced landslide and debris flow hazard chain. *Front. Earth Sci.* 11:1157881. doi: 10.3389/feart.2023.1157881

An integral assessment of landslide dams generated by the occurrence of rainfall-induced landslide and debris flow hazard chain

Laura Ortiz-Giraldo*, Blanca A. Botero and Johnny Vega

Facultad de Ingeniería, Universidad de Medellín, Medellín, Colombia

Landslides caused by rainfall are one of the most frequent causes of disasters in tropical countries and mountainous terrain and can block rivers generating landslide dams. This paper presents a methodology for the estimation of the obstruction of water streams generated by rainfall-induced shallow landslides. The spatial distribution of the landslide hazard was estimated in terms of the Factor of Safety (FoS) values using the deterministic method with physical basis SLIDE (Slope - Infiltration - Distributed Equilibrium). The rainfall regimes of the study area were estimated by means of a simple scaling Log Normal Model. Subsequently, the resulting areas with a high hazard level that could detach and reach the riverbed were identified as sources for the simulation of the debris flow runout using the Rapid Mass Movement Simulation model with its debris flow module, (i.e., RAMMS-DF), estimating zones of the riverbed that should be analyzed in detail. Finally, the effects of river channel obstructions generated after debris flow movement were analyzed by means of the Iber, a well-known, physically based 2D hydraulic model and their possible changes on the river hydraulic. In order to generate a workflow that allows the application of the SLIDE methodology and the preparation of inputs for the subsequent processes of debris flow propagation and hydraulic modeling of the river corridor of analysis, a Python-based toolbox was created. Our results highlight the changes in the fluvial dynamics in the corridor of the river of analysis after the landslide dams generated by the occurrence of rainfall-induced landslide and debris flow hazard chain for the different return periods. In all cases, the material deposited in the river channel was sufficient to change the hydraulic regime of the river corridor, showing longer delay times in the transit of the flow, in addition to the decrease in the specific flow. This would imply a water shortage in the study basin of the hydroelectric project; however, in the scope of this project it is not possible to really determine the real effects that could be generated by this event.

KEYWORDS

landslide dam, debris flow, hazard chain, rainfall-induced landslide, RAMMS model, SLIDE model

1 Introduction

Landslides triggered by rainfall are one of the most frequent causes of disasters in tropical countries and mountainous terrains (Cui and Guo, 2021; Aristizábal et al., 2022; Dhanai et al., 2022). Climate change has led to an unprecedented increase in the frequency and intensity of rainfall, and a greater number of landslides are triggered by rainfall along hillsides. (Thomas et al., 2023). According to the World Health Organization, between 1998 and 2017, landslides affected an estimated 4.8 million people and caused more than 18,000 deaths worldwide (UNISDR and CRED, 2017). Landslides can be classified into different types based on the slope movement mechanism and rate, the material involved, the mechanical behavior, or the movement stage (Hungr et al., 2014). Debris flows are a landslide type (or a specific case), composed of several materials including soil, rock, and water. It is an extremely rapid flow-type landslide, which tends to travel long distances from its source in a steep channel (Arghya et al., 2022; Trujillo-Vela et al., 2022). Landslides and debris flows are frequent hazards that seriously endanger lives and livelihoods, causing fatalities, property damage and impacts on engineering projects. They are a cause for concern worldwide (Gan and Zhang, 2019; Baggio et al., 2021; Fustos-Toribio et al., 2021).

Rainfall-induced shallow landslide has the potential to evolve into debris flows and thereby gather more channel sediments and cause significant destruction along its path and in the downstream areas (Thomas et al., 2023). Debris flows can block rivers and form hazards, causing great losses, i.e., they correspond to a chain of hazards caused by landslides and rainfall-triggered debris flows (Chen et al., 2022). A chain of multiple hazards can be defined as a series of hazards that occur in succession that are triggered by a single hazard resulting in an amplification of damage in both time and space relative to each of the constituent hazards if they happened separately (Zhu et al., 2021; Guo J. et al., 2022). These natural hazards are often caused by the diverse geographic and physiographic features, triggered by both natural as well as anthropogenic factors. Alike, blockage of a mainstream by tributary debris flow events is a natural phenomenon of local river evolution between the confluences of the streams (Vega and Hidalgo, 2021).

The physically based method is a respectable approach for susceptibility analysis of shallow landslides because is the most physically (geotechnical) simple model, based on the limit equilibrium theory, which is used to ascertain a given slope's safety factor, reproducing the physical processes governing landslide occurrence (Chae et al., 2017; Armaş et al., 2021). Some examples are The Transient Rainfall Infiltration and Grid-based Regional Slope Stability (TRIGRS); Shallow Landslide Stability Model (SHALSTAB); The Fast Shallow Landslide Assessment Model (FSLAM); Probabilistic estimation of the threat due to landslides and mass movements in hillsides (EPADYM) and the model Slope Infiltration Distributed Equilibrium (SLIDE), among others. These models evaluate the spatial distribution of the hazard in terms of Factor of Safety (FoS) for rainfall-induced landslides (Chikalamo, 2018), applying infinite slope models which make them mostly applicable for the analysis of shallow landslides. The TRIGRS model considers partially saturated conditions and transient flows (Aristizábal et al., 2022; Rana and Babu, 2022); SHALSTAB model assumes that local surface topography is the dominant control of landslide occurrence, which makes it appealing for DEM-based landslide analyses (Moreira Melo et al., 2021); FSLAM was developed to calculate large areas (>100 km^2) with a high-resolution topography in a very short computational time (Medina et al., 2021). This model includes two different sub-models: the hydrological model, which calculated the rainfall infiltration into the soil layer and the resulting position of the water table, and the geotechnical model, which applies the widely used infinite slope theory to calculate the slope stability (Guo Z. et al., 2022; Hürlimann et al., 2022); EPADYM model calculates the reliability index, failure probability and stability factor of a slope under seismic and static conditions (Hidalgo and Vega, 2021). On the other hand, SLIDES illustrates the dynamical relations between FoS and accumulation of rainfall over time when the shallow surface of soil reaches saturation and has enormous potential for rainstorm-induced landslide evaluation (Chikalamo, 2018; Chen et al., 2019).

Debris flow runout may refer to the entire horizontal distance (path) length measured on a map. Determining this distance of the landslide-debris flow chain is one of the most important tasks during the hazard assessment and zoning because of the complexity of the phenomenon and the variability of controlling factors (Liu et al., 2022). With the development of computation, numerical methods have gradually become an efficient and powerful tool for simulating landslide movement processes and understanding the complex dynamic process (Zhang J. et al., 2022). Some examples of debris flow propagation prediction models that have recently been widely used in the literature are: The FLO 2D model is a 2D finite difference model that can simulate non-Newtonian floods and debris flows (Do et al., 2020; Liu et al., 2022); the Flow path assessment of gravitational hazards at a regional scale model (FLOW R) can calculate the extent and paths of debris flows (Franco-Ramos et al., 2020; Arghya et al., 2022); the Debris flow simulation 2D model (DFS 2D) can be used to simulate the dynamics of debris flows considering multiple rheologies (Abraham et al., 2022; Liu et al., 2022). The Rapid Mass Movement Simulation model (RAMMS) has a debris flow module to analyze the dynamic process of geological disasters (Graf and McArdell, 2008; Christen et al., 2010; Franco-Ramos et al., 2020; Mikoš and Bezak, 2021; Zhang X. et al., 2022; Liu et al., 2022; Zhou et al., 2022); RAMMS is one of the most used one's landslide runout methods allowing estimate flow velocity, deposition height, accumulation distribution, flow velocity and debris flow development propagation footprint of the debris flow materials (Mikoš and Bezak, 2021; Zhang X. et al., 2022; WSL Institute for Snow and Avalanche Research SLF, 2022).

Some authors have already made successful progress in landslide-debris flow chain evaluations, as (an et al. (2021); Zhou et al., 2022 and (Zhang Y. et al., 2022). However, most of them are focused on the analysis of the chain up to debris flow runout modeling. One of the current challenges is to improve the predictability of such events reaching a river and generating landslide dams, causing the loss of water supply in communities and hydroelectric projects. River-blocking induced debris flows have become common in numerous catchments of mountainous topography with a strong hydroclimatic variability, in response to climate and environmental changes (Qiang et al., 2020; Zhou et al., 2022). The early identification of landslide dam disasters has

received considerable attention in recent years. Extensive studies are available nowadays (Kun-Ting et al., 2018; Li et al., 2020a; Hu et al., 2020; Zhang Y. et al., 2022). These studies have focused mainly on the blockage of the river, classifying it as a partial or total obstruction, as well as the evaluation of the interaction of the landslide dam material with the river flow.

Hydraulic modeling has been used to evaluate changes in flow behavior (González-Cao et al., 2019). For hydraulic modeling, there is extensive use of hydraulic simulations in the literature of the HEC RAS 2D and Iber models. HEC RAS 2D is a model for unsteady flow based on the Navier–Stokes (Roldán et al., 2022); Iber 2D is a free model package for simulating unsteady free surface turbulent flow and transport processes in shallow water flows based on 2D Saint-Venant Equations (Bladé et al., 2014; Hafnaoui and Debabeche, 2021). Iber represents a better option for modeling rivers with more irregular geometries, especially when there is a special interest in the hydraulic behavior of the flow it has been successfully used in a wide range of applications, including impacts of dam breakage (Hoyos and Botero, 2019; Ruiz-Villanueva et al., 2020; Díaz-Salas et al., 2021; Pérez-Montiel et al., 2022).

The early identification of landslide dam disasters has received considerable attention in recent years, these studies have mostly focused on dam failures and dam-breach floods (Nian et al., 2021; Zeng et al., 2022). Nevertheless, little is known about water shortages after a river blockage due rainfall-triggered landslides, nor about the change in the flow regime after a landslide dam formation. This leads to the need to evaluate the possible effects of these disaster chains on water resources and adopt strategies to protect the water streams of supply. It is fundamental to investigate and complement those studies of disaster chains up to the evaluation of the generation of the obstruction and its possible impacts. Uncontrolled transport of debris flow can be a major problem in maintaining the proper operation of hydroelectric facilities in the future (Walczak et al., 2021). Enhanced resilience to future obstructions on the river generated by debris flow landslide dam chain triggered by rain will be essential for reliable electricity supply and ensuring greater long-term opportunities (Chen et al., 2023).

In this paper, we propose an integral assessment of river channel obstructions due to the occurrence of the chain of hazards produced by rainfall-induced landslides and the subsequent occurrence of debris flows in tropical mountainous regions. We estimated the spatial distribution of shallow landslides hazard in terms of FoS values by the physical-based model SLIDE. High hazard zones that may detach and generate obstructions in a river channel were identified. Debris flow propagation was carried out using the RAMMS model. Then, an analysis of the consequences considering the deposition height on the riverbed was conducted, considering its influence on river hydraulics and potential water supply affectation. Changes in the hydraulic variables of the river flow after the debris flow movement and the generation of landslide dams in the river channel were analyzed using the Iber model, identifying their possible influence on the water shortage in the intake structures of the hydroelectric project. The methodology was applied in a sub-basin river for hydropower generation purposes. A toolbox was created with the purpose of generating a workflow that allows the application of the SLIDE methodology and the preparation of inputs for the subsequent processes of debris flow propagation simulation and hydraulic modeling of the fluvial corridor of analysis. This toolbox it is fully integrated with one of the most popular GIS environments, ArcGIS, developed by the company Esri (Dysarz, 2018).

The proposed methodology can be a useful tool for specialists interested in taking mitigation measures that contribute to the reduction of economic losses, development of protocols in the event of a disaster or choosing possible locations for the construction of new infrastructures considering exclusion criteria, as well as for its implementation in the updating and future land planning for local authorities. The understanding of the complete process of the hazard chain can be a key tool to reduce these natural disasters.

2 Study area and data sources

The study area is located in the northwest of Colombia, northern part of the Central Andes Mountain range, as shown in Figures 1A, B, the water intake structure for the Small Hydropower Plant (SHP) defines the subbasin of study in the Ovejas river. The subbasin area of the Ovejas River is 77 km^2 and it has an average flow of 3.2 m^3/s, and a complex topography formed by a mountainous relief with altitude range is between 2,198 m.a.s.l. and 2,600 m.a.s.l. Precipitation is mainly convective and orographic type with a multiannual average precipitation of approximately 2,200 mm/year. Its climate corresponds to a subtropical humid forest and humid tropical hydrometeorological conditions (Aristizábal et al., 2017). The rainfall in the study zone shows a bimodal distribution peaking between April–May and October (García-Delgado et al., 2021). The available meteorological stations represent the pluviometric phenomena in the study area. This information was acquired through the National Meteorological Service (IDEAM).

The study basin was traced at the water intake Structure for the SHP (Figure 1C) based on an Advanced Land Observing Satellite Digital Elevation Model (ALOS DEM) downloaded from the Alaska Satellite Facility (ASF DAAC, 2015 and PALSAR_Radiometric_Terrain_Corrected_high_res,), with a grid cell size of 12.5 m and with this information the morphometric characteristics and hydrologic parameters of the study basin were estimated. (Figure 1C). Additionally, a river corridor of analysis of 40 m on each side of the river and a longitudinal distance of 600 m upstream from the water intake structure site is proposed to model the runout of rainfall induced landslide debris flow hazard chain and hydraulic processes in such a way that it would be possible to develop the project study in feasible computational times (Figure 1D). For this river corridor was used the detailed topographic information provided by the SHP project owner. This topography corresponds to a LiDAR survey (Light Detection and Ranging or Laser Imaging Detection and Ranging), acquired in 2018, with a pixel resolution of 0.5 m. This information was complemented with information of the study area obtained from the River Basin Regulation and Management Plans (AMVAÁrea Metropolitana del Valle de Aburrá, 2015).

The basin is characterized as a topographically undulating area, with steep slopes in some areas of the upper part of the river corridor of analysis and in the immediate of the SHP water intake structure, reaching up to 67% slope (Figure 2A). These topographical factors condition climate and influence the spatial rainfall distribution and

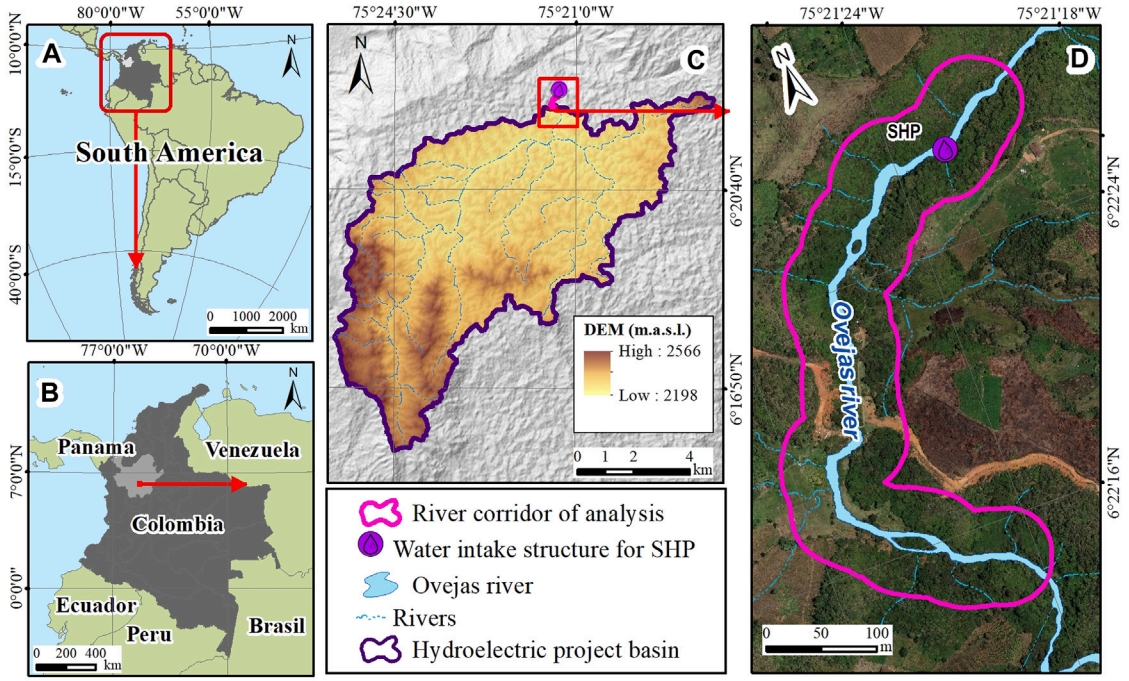


FIGURE 1
Location map. **(A)** Location of Colombia on a continent scale. **(B)** Location of Antioquia on a country scale **(C)** Location of the hydropower basin with a digital elevation model. **(D)** River corridor of analysis.

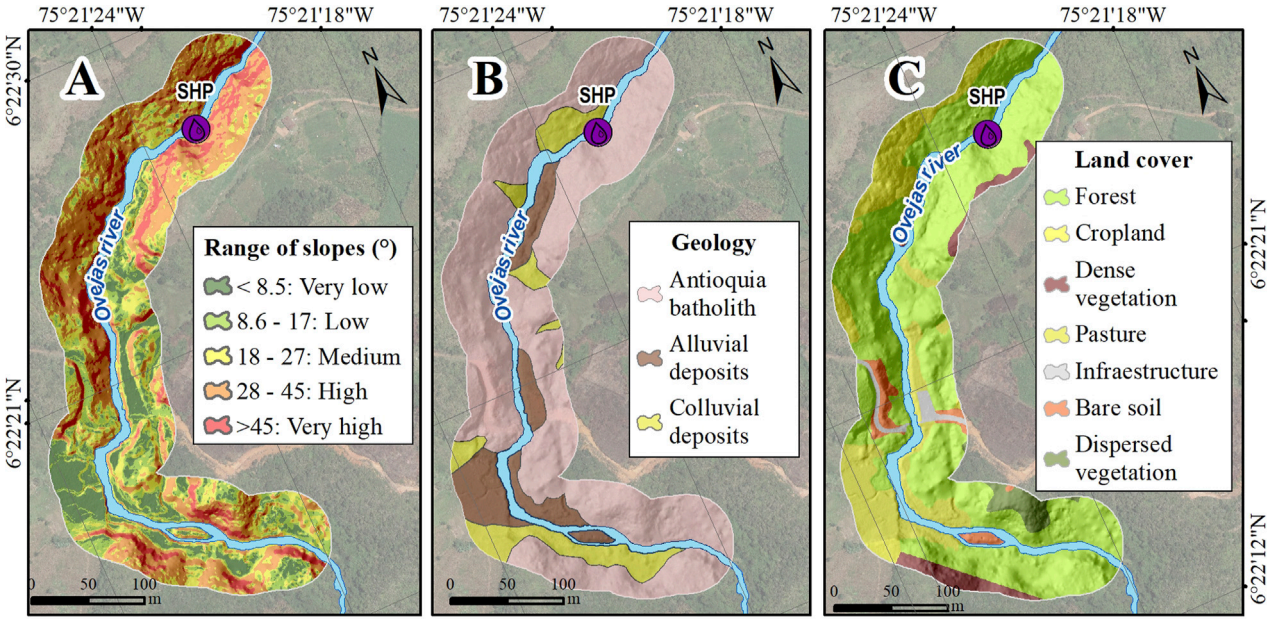


FIGURE 2
(A) Slope map of river corridor of analysis. **(B)** Spatial distribution of the geological units. **(C)** Land coverage map.

create favorable conditions for the formation of landslides and debris flows.

The river corridor of analysis comprises three main geological units: Antioquia batholith, Alluvial deposits and Colluvial deposits (Integral, 2020). The riverbed of the Ovejas river is narrow and contains alluvial deposits with a predominance of fine granulometry and large blocks of fresh rock from lateral landslides in the valley (Figure 2B) (Integral, 2020). The Antioquia batholith is the rock that

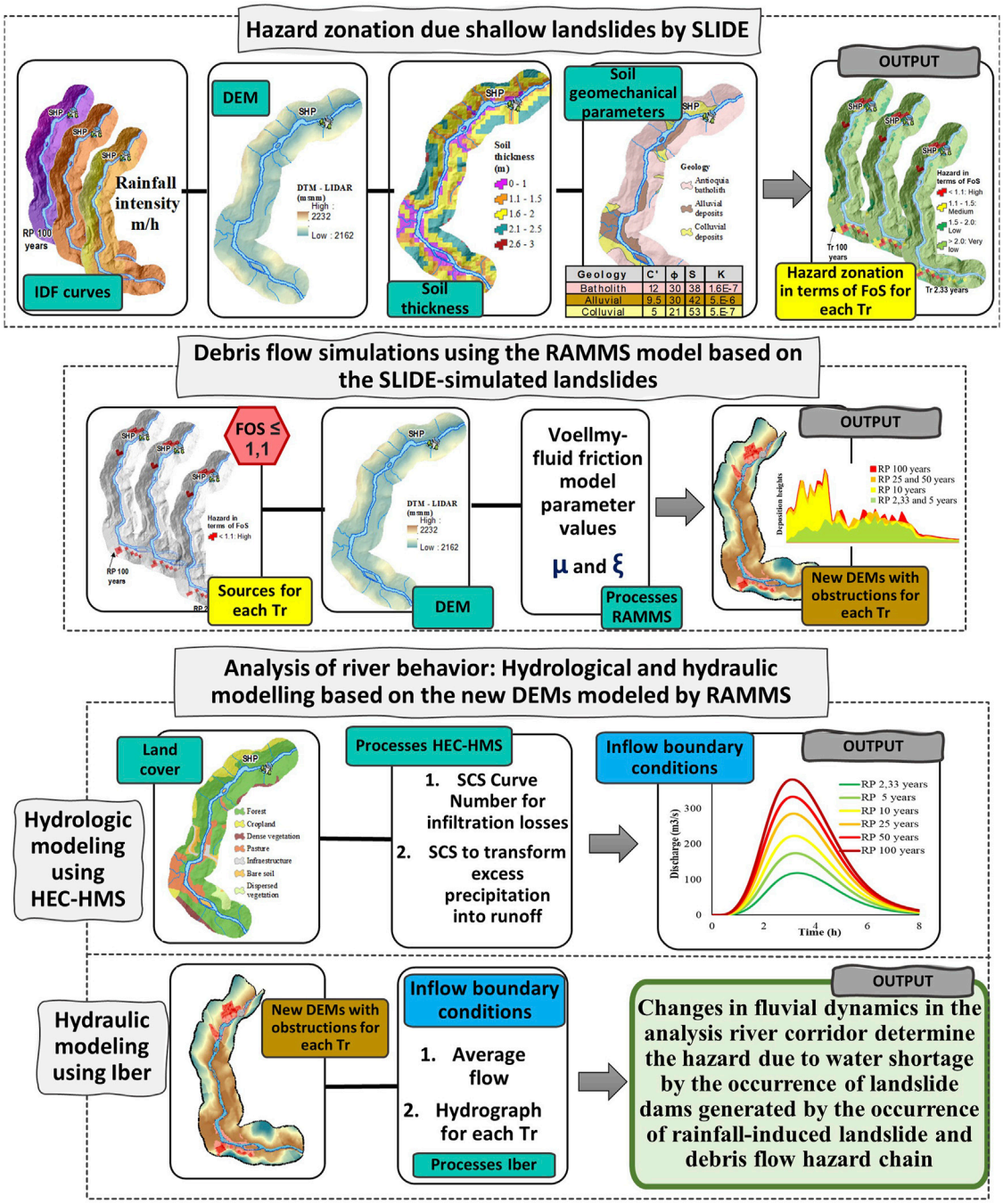


FIGURE 3
Workflow for the proposed methodology in this study.

outcrops in the study area, is a massive intrusive body of predominantly sandy silt granulometry (Marin et al., 2021). There are several colluvial deposits on the left bank of the Ovejas river, they are the result of current erosional processes. In general, they show high humidity and reptation phenomena. The alluvial deposits are poorly selected deposits with variable granulometry from gravel to silt. The geological map and its information of soil and lithology of the area of study were obtained from the detailed prefeasibility studies provided by the SHP project owner (Integral, 2020). The land cover (Figure 2C) corresponds mostly to natural forest, but there are also some areas of crops, bare soil, vegetation, and infrastructure zones.

3 Materials and methods

3.1 Methodology

In this study, the methodology can be schematized into three successive phases: The workflow is presented in Figure 3. The first

implemented process was the landslide hazard assessment using SLIDE model. This method was developed by Liao et al. (2010) to evaluate the spatial distribution of the hazard in terms of FoS for rainfall-induced shallow landslides. As inputs, the model requires IDF (intensity-duration-frequency) curves, soil thickness and soil geomechanical parameters. The second process was the modeling of debris flow propagation (runout) using the Rapid Mass Movement Simulation model with its debris flow module (RAMMS- DF). Inputs for RAMMS were the DEM and the release areas resulting from the SLIDE hazard zoning methodology. The rainfall-runoff processes were simulated by HEC-HMS, the Soil Conservation Service-Curve number (SCS-CN), SCS Unit Hydrograph and Muskingum methods were used for loss, transformation, and routing calculations, respectively, obtaining the hydrographs for different return periods (Tr). The last process was the simulation of the effect of the river obstruction in the river corridor using the Iber model, which requires as input the flow hydrographs for the return periods analyzed and the heights of material deposition obtained from the RAMMS simulation.

This methodology was developed from the semi-automation of the analysis processes, through the generation of a toolbox for ArcGis based on scripting language Python 2.7 based on the ArcPy package with the purpose to generate a workflow that allows the application of the SLIDE methodology and the preparation of inputs for the subsequent processes of simulation of debris flow propagation and hydraulic modeling of the river corridor of analysis. This optimizes and complements the work and decision making of the expert for the application of the methodology from a set of procedures, tools and utilities specially designed to process data under the environment of Geographic Information Systems (GIS). For the use of the toolbox is necessary to have the GIS raster cells of the digital elevation model, the geological and geomorphological information of the study area, coverages, soil parameters, rainfall distribution as a triggering factor, among other input parameters of the different models. At forward subsections these phases are explained in more detail.

3.2 Shallow landslide assessment using SLIDE model

The spatial distribution of rainfall induced landslide hazard in terms of FoS values was estimated by a deterministic model with physical basis SLIDE, which integrates the infinite slope model with a rainfall infiltration model. The analysis was focused on shallow landslides to identify those areas susceptible according to the intrinsic soil parameters and the morphometric configuration of the terrain. Additionally, the model allows to simulate rainfall events as a triggering factor and the response of soil to such events. The FoS was obtained by the ratio of stabilizing forces to destabilizing forces. Slopes are typically considered stable when FoS is ≥1; while a landslide is predicted when FoS <1 (Liao et al., 2012). Thus, the transition between stability and collapse can be seen mathematically as a decrease in the FoS to values below unity (Montoya Botero, 2018; Tyagi et al., 2022). In this study, high hazard zones were considered with a FoS ≤1.1, thus presenting a conservative scenario (Vega and Hidalgo, 2016).

Given the complexity of shallow landslide studies, the following simplifying assumptions are made in the SLIDE model: The slope is considered infinite with low soil thickness, the failure is considered flat and water downflow occurs, both parallel to the slope and the evapotranspiration, underground seepage and surface flow are avoided in the water balance. Therefore, the infiltration amount approximately equals to the maximum of saturated hydraulic conductivity (Li et al., 2020b). This model considers some simplified hypotheses on the water down flow and defines a direct correlation between the FoS of the slopes and the rainfall depth (Montrasio et al., 2013). A relation between the rainfall amount and the final expression of FoS has been set up and translated into a simple mathematical formulation of model SLIDE. Γ gamma is the variable that introduces in a compact way, the variables that represent the soil properties: specific gravity, porosity, degree of saturation, while Ω Omega involves the geometric variables such as slope and soil thickness susceptible to sliding. The FoS is calculated as follows (Liao et al., 2011):

$$FoS = \frac{\cot\beta.\tan\phi.[\Gamma + m_t.(n_w - 1)] + C.\Omega}{\Gamma + m.n_w} \quad (1)$$

$$\Gamma = G_s.(1-n) + n.S_r \quad (2)$$

$$\Omega = \frac{2}{\sin 2\beta.H.\gamma_w} \quad (3)$$

$$n_w = n(1 - S_r) \quad (4)$$

Where β is the slope angle [°], ϕ is the friction angle [°], C is the soil cohesion [kpa], is a fractional parameter between 0 and 1 of soil thickness by infiltration. Parameter m, which changes in time, calculated by using Eq. 5. G_s is the specific gravity, n is the porosity, S_r is the degree of saturation [%], H is the soil thickness [m] and γ_w is the unit weight of the water [kN/m^3] (Montrasio et al., 2013; Zhang J. et al., 2022).

The saturated part of the soil layer that directly correlates precipitation with FoS is expressed by the parameter m_t, which is a dimensionless thickness and during numerical simulation must be calculated for each time step (Marin et al., 2021). The initial value of m could be determined by an *in-situ* test of the water table or from the assumption of this value according to the knowledge of the study area. O_t represents the water outlet of a finite portion of a slope of finite length and is calculated by Eq. 5. (Liao et al., 2010).

$$\begin{cases} m_1 = 0 \\ O_t = K\sin\beta\ .\ m_t.\ H.\cos\beta\ .\ \Delta t \\ \Delta m_t = \dfrac{(I_t - O_t)}{n\ .\ H.(1 - S_r)} \\ m_{t+1} = m_t + \Delta m_t \end{cases} \quad (5)$$

Where t is time [h], Δt is time step m_1, is initial value of m and mt is calculated at each time-step. I_t is rainfall intensity [m/h] and K is the hydraulic conductivity of saturated soil [m/h] (Liao et al., 2010). m was supposed to be 0.3 in the river corridor of analysis.

The physical mechanism of movement corresponds to superficial, translational landslides with rupture surfaces parallel to the slope surface. This model makes large assumptions for the estimation of the FoS hazard, so it was considered to resample the digital elevation model of 0.5 m lidar resolution of the study area in an elevation model with a spatial resolution of 10 m in the modeling

of Landslide Hazard assessment in terms of FoS. The zoning was performed for the entire study basin; however, to carry out the study in computationally feasible terms, the analysis river corridor (Figure 1D) was delimited for subsequent debris flow runout processes and dynamic evaluations of the hydraulic behavior of the river.

3.2.1 Rainfall intensity

The spatial distribution, duration, and intensity of rainfall play an important role in triggering landslides (Liao et al., 2011). The rainfall scenarios in the slope stability model SLIDE are represented by the rainfall intensity values (Eq. 5 in Eq. 2). There were no meteorological stations with available IDF curves in the study area, so these curves had to be estimated using the simple scaling methodology (Pulgarín and Poveda, 2008). The expression corresponding to the simple scaling is shown in Eq. 6:

$$I_{d,q} = \frac{E\left[I_{dref}\right]^2}{\sqrt{E\left[I_{dref}\right]^2}} exp\left[\phi_q . \sqrt{ln\frac{E\left(I_{dref}\right)^2}{E\left[I_{dref}\right]^2}}\right]\left(\frac{d}{dref}\right)^{\theta} \quad (6)$$

Where $I_{d,\,q}$ is the maximum intensity for a duration d for a non-exceedance probability [m/h], I_{dref} maximum intensity of a reference duration [m/h] $E[I_{dref}]$ expected value [m/h], ϕ_q quantile for a non-exceedance probability q from the standard Normal distribution [m/h], θ the scaling exponent.

The reference duration was 24 h according to the information available in IDEAM rainfall gauges. In the absence of an annual series of maximum rainfall in 24 h for the study basin, the annual series of maximum daily rainfall available were used, considering an increase of 11% (Pulgarín Dávila, 2009). For the analysis river corridor, hourly rainfall intensities were estimated up to the time of concentration of the SHP basin for the analyzed return periods of 2.33, 5, 10, 10, 10, 25, 50, and 100 years.

3.2.2 Soil thickness

The depth of the potentially unstable soil is a critical parameter that strongly influences the stability of slopes (Oliveira et al., 2017). The evaluation of such thickness, susceptible to sliding or detachment, is essential before the analysis of problems involving hydrological processes developed on slopes, as well as slope stability (Fernandes Azevedo et al., 2015). The estimation of soil thickness is based on the behavior of the soil in relation to the slope of the terrain; high slopes imply less soil accumulation. The model of the soil susceptible to sliding is represented by the following equation (Montoya Botero, 2018).

$$H = (h_{max} - h_{min})\left\{\left[1 - \frac{1}{tan(\theta_{lim})} tan\,\theta\right][tanh(ax)]\right\} + h_{min} \quad (7)$$

Where H Soil thickness [m]; h_{max} Maximum soil thickness [m]; h_{min} minimum soil thickness [m]; θ slope angle value [°]; θ_{lim} is the slope angle value threshold for which higher values imply that the thickness of soil susceptible to sliding is minimal/negligible or even null [°]; a Dimensionless parameter that controls the estimating of soil thickness depending on the curvature of the terrain; the values were chosen according to predominant soil types of the geological units in tropical zones. The curvature controls the zones of water accumulation or spreading, which directly influences the degree of soil moisture and/or saturation and, therefore, the generation of debris flows (Náquira Bazán, 2009), and x represents the horizontal distance from a cell to the nearest drainage network. The term $tanh(ax)$ corresponds to the calculation of the hyperbolic tangent of ax, generating an interaction that acts in such a way that the greater the horizontal distance from a given point to the nearest drainage, the greater the growth tendency of the soil layer (Montoya Botero, 2018).

Slope steepness is one of the key factors in determining landslide debris flow hazard chain. Most avalanches occur on slopes that have an incline of 35°–45° (Calcaterra et al., 2022). The choice of the slope limit in this study was made after analyzing the descriptive statistics of the slope distribution in the analyzed river corridor (Figure 2). The mean slope angle is 24°, while the standard deviation is 11°. Thus, it was decided to consider a critical value (θ_{lim}) for the slope of 35°, slightly higher than the measure comprising the mean plus standard deviation. The value of the variable a was set at 0.04 (Montrasio et al., 2013), while the maximum and minimum predicted thicknesses were assigned for each geologic unit, for the potential depth of rainfall-induced landslides (Integral, 2020). By applying a map algebra through the ArcGis platform as a Geographic Information System (GIS), the parameters involved were integrated and the model of soil thickness susceptible to landslide was spatialized.

3.2.3 Model parameters

For the application of the SLIDE model, the LiDAR topography survey (0.5 m pixel resolution) was resampled to obtain a DEM with a spatial resolution of 10 m, because the physical mechanism of movement corresponds to shallow landslide with failure surface parallel to the slope surface. For the geological characterization of the study area, the available information from the Colombian Geological Service was reviewed, such as topographic maps, geological maps and aerial photographs. Additionally, the owner of the SHP project carried out a soil exploration of the study site and subsequent laboratory analysis to determine the lithological characteristics, the thickness of the meteorization profiles, as well as the morphometric and morphodynamic configuration of the terrain. According to these exploration studies and the correlations reported in the literature for the soils found in the study area, the strength parameters and the geomechanical properties of the soils were defined, as well as the adopted depth values for the estimation of the soil thickness susceptible to landslide. Figure 2B shows the spatial distribution of the geological units of the study basin, and the soil properties required for SLIDE model and values for soil thickness estimation are summarized in Table 1.

3.3 Landslide propagation (runout) modeling using RAMMS

Rapid Mass Movement Simulation (RAMMS) with the debris flow module (called RAMMS: DEBRIS FLOW or RAMMS-DF) was developed to simulate the runout of muddy and debris-laden flows in complex terrain (Mikoš and Bezak, 2021) for practical and research-oriented applications. The model used in RAMMS employs the Voellmy–Salm fluid model, is based on a depth-

TABLE 1 Soil parameters for landslides assessment.

	C′ (kPa)	φ (°)	Sr (%)	K (m/s)	h min (m)	h max (m)
Antioquia batholith	12.0	30	38.49	1.60E-07	1.2	6
Alluvial deposits	9.5	30	41.65	5.00E-06	1.0	6
Colluvial deposits	5.0	21	52.07	5.00E-07	1.0	4

averaged 2D solution to the implemented law of motion over 3D topography. The core of the program is an efficient second-order numerical solution of the depth-averaged equations of motion (the shallow water equations) for granular flows (Liu et al., 2021). This model can quickly simulate debris flows, and obtain motion parameters including motion distance, velocity impact pressure, and flow path, in three-dimensional terrain (Zhang X. et al., 2022).

Debris flow comprises three distinct zones: the source zone, the transportation zone, and the deposition zone. The area from where the debris flow initiates or the mobilization starts is the source zone (Dash et al., 2023). The initial height is determined by the user when defining the source area of the debris flow as a single or multiple block-release (initiation) zones can easily be specified by polygons by GIS-type drawing tools; alternatively, an input hydrograph can be used to specify the discharge as a function of time (Zimmermann et al., 2020).

The landslides footprint of the sliding mass emerging from was estimated using RAMMS-DF. According to the available information in the study basin the block release methodology was chosen using a polygon layer of triggering zones. The release areas were defined by polygons with shapefile containing the results of the SLIDE methodology hazard zoning methodology with high hazard zones. (Section 3.2). Source areas were considered as those with FoS <1.1 (Hidalgo and Vega, 2021). The block release module needs to use a release area with a given initial depth, which was set with the soil thickness estimated previously (Section 3.2.2). Additionally, for the launch of RAMMS module a DEM is necessary, hence, the LiDAR data was used.

There are several equations that control the modeling of RAMMS model, in addition to their corresponding input parameters, highlighting some of them such as the flow height (H), the gravity (g), slope angle (β), and the Voellmy friction coefficients: dry Coulomb-type friction μ and turbulent viscous friction ξ, among others (RAMMS, 2022). Friction coefficients μ and ξ, which are constant given by user, determine the flow properties of debris flow in RAMMS (Roldán et al., 2022). The parameter ξ is dominant for debris flow in quick flowing state and μ is dominant for debris flow in almost static state. These two parameters are usually calibrated, although other parameters such as stop parameter or simulation resolution also influence the modelling results (Mikoš and Bezak, 2021; Calcaterra et al., 2022). For more information about RAMMS model and calibration parameters, (ramms.slf.ch/en/modules/debrisflow.html).

For the river corridor of analysis, different scenarios were modeled by varying dry Coulomb friction coefficient μ and a turbulent friction coefficient ξ (m/s^2), to represent different scenarios to obtain the maximum and minimum flow propagation, observing how the magnitude of the debris flow decreases and increases slightly as μ and ξ increase. Values of μ typically range from 0.05 to 0.4 (dimensionless). Values of μ greater than 0.4 rarely provide useful simulation results. ξ describes the turbulent behavior of the flow. Small values of ξ are usually reported for granular flows, while relatively large values of ξ are sometimes associated with muddy flows. ξ is more delicate to calibrate and the tool suggests values between 100 and 200 m/s^2 (WSL Institute for Snow and Avalanche Research SLF, 2022). Nevertheless, since there are no records of historical events in the study area that provide information that completely describes the rheological parameters of the soil, the Voellmy friction parameters were estimated according to the known characteristics of the soil mentioned in Section 3.2.3 and the suggestions of the RAMMS user's manual, obtaining μ = 0.4 and ξ = 100 m/s^2.

3.4 Hydrological and hydraulic modelling of the main water stream

The HEC-HMS model was used to analyze the rainfall–runoff processes of the SHP basin and to compute the hydraulic processes in the river corridor of study the numerical model Iber was implemented.

3.4.1 Hydrologic modeling using HEC-HMS

The Hydrologic Engineering Center and Hydrologic Modeling System HEC-HMS (Feldman, 2000 and U.S. Army Corps of Engineers, 2018), could be used for many hydrological simulations (Hamdan et al., 2021)). HEC-HMS is an accessible rainfall-runoff model, being one of the most widely used for hydrological simulations and offers a wide range of methods and modules to estimate runoff from rainfall, e.g., loss models, transformation models, routing methods, among others (Sahu et al., 2020; Reddy and Pramada, 2022). For more details about this model and its different methods, please consult in US Army Corps of Engineers Hydrologic Engineering Center (www.hec.usace.army.mil/software/hec-hms/).

The hydrologic model HEC-HMS was employed to determine the peak discharge for each return period for the study basin because in the study area there are no discharge records series to perform a frequency analysis and obtain the maximum discharge for the return periods of the study. This approach employs a rainfall-runoff model. The hypothesis that supports this methodology establishes that a maximum rainfall estimated for a given return period produces a maximum discharge for that same return period. Three main processes were considered in the methodology developed in this case of study:

- The Soil Conservation Service (SCS) Curve Number method was used to compute the infiltration losses. This method

implements the curve number methodology for incremental losses since it was designed to calculate the infiltration during periods of heavy rainfall (González-Cao et al., 2019). Based on land cover and soil texture a CN of 81.5 was obtained, which suggests that the general conditions of the territory tend to favor runoff over infiltration. The input parameters were the vegetation cover maps corresponding to antecedent moisture condition II. The soils of the study watershed are part of hydrological group C.

- The SCS UH (Unit Hydrograph) method was applied to transform the excess rainfall into runoff, having as input the maximum rainfall for each return period with rainfall duration equal to the time of concentration of the study watershed.
- The Muskingum routing method was used for flow routing of the channel reaches because it provides a good approach in basins with similar slopes (González-Cao et al., 2019; Hamdan et al., 2021).

3.4.2 Hydraulic modeling using iber

Iber is a free software that implements a family of bidimensional hydraulic mathematical models for the simulation of free surface flows in rivers and estuaries (Bladé et al., 2014; Pérez-Montiel et al., 2022). The equations are solved by the finite volume method on the grid nodes. The numerical schemes used in Iber are robust and stable in any situation, suitable for discontinuous flows, specifically for torrential channels and irregular regimes. The finite volume numerical methodology solves Shallow Water Equations in two dimensions (García-Alén et al., 2022). Modeling in Iber model requires a three-dimensional terrain model or mesh as starting data and the incorporation of the boundary conditions, the initial conditions, as well as the assignment of the terrain roughness coefficients (García et al., 2022). More information about this model can be found at (www.iberaula.es).

To construct the geometry of the meshes for each scenario, the information from the Lidar survey of the study area processed in ASCII format was used, in addition to the new DEMS modified with the heights of the landslide dams obtained with the RAMMS model in Section 3.3, building models with a maximum size of 1 m, minimum size of 0.5 m and tolerance of 0.1. Studies on the influence of mesh size in hydraulic modeling in flood plains with elements from 6 to 24 cm have given satisfactory results in flood levels (Bomers, 2020).The mesh size used for Iber is smaller for a better detail of the results but at the same time considering a feasible computation time, so it was considered adequate.

Roughness coefficients expressed by Manning's n values were assigned from the land cover map (Figure 2C). Two types of inflow boundary conditions were established: the estimated hydrographs for each return period and the average flow of the stream. For the outflow conditions, given that there is no information associated with flood events such as depths at the end of the analysis section, a supercritical/critical regime was assumed for each scenario (Bladé et al., 2014). Once the simulation was completed, the results were exported to raster for better handling of the information in ArcGIS (www.esri.com).

4 Results

4.1 Shallow landslide assessment

The rainfall intensity for each return period was estimated with the generated IDF curves. Figure 4 shows the rainfall intensity maps with a duration equal to the estimated basin time of concentration of

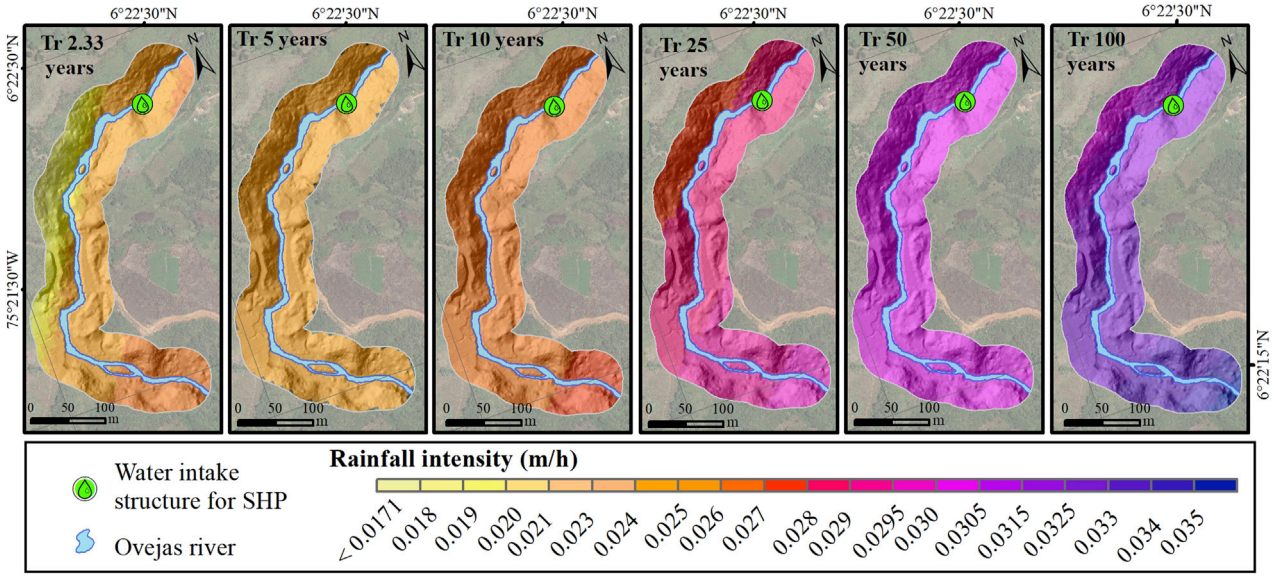


FIGURE 4
Rainfall intensities for the basin concentration time.

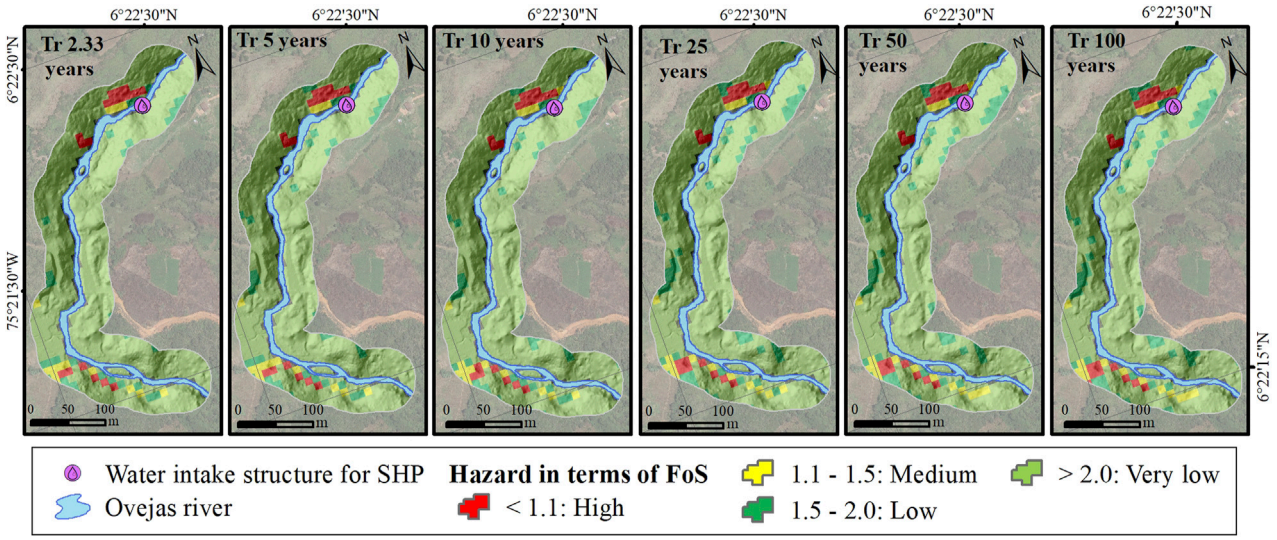


FIGURE 5
Landslide Hazard assessment in terms of FoS.

3 h. The highest intensities were found towards the south and southeast of the river corridor of analysis.

From the rainfall intensities for each return period, SLIDE model estimated the landslide hazard according to the FoS values. Four categories based on FoS values were used. Lowers values indicates higher potential for landslides occurrence. (Figure 5).

Based on Figure 5, it is observed that moderate and high hazard zones were distributed mainly in soils susceptible to landslides of greater soil thickness, where higher rainfall events progressively affect soil saturation and its possible failure, highlighting the influence of topographic features on the stability of the terrain. It is important to highlight that although slopes are a factor that have a high degree of influence, they do not control the landslide hazard zoning and it is affected by soil parameters, including the soil thickness susceptible to sliding. Since the model follows a correlation between FoS and rainfall, the high and medium hazard zones determined by soil thicknesses with very low FoS (H < 1.6 m), increase simultaneously with the return period analyzed for each scenario (with respect to the very low hazard zones). However, it should be noted that for the return periods of 2.33 and 5 years, the safety factor had the same spatial distribution. The same consideration is true for the 25 and 50 years return periods.

The results of the landslide hazard assessment in terms of FoS and landslide propagation were compared with the historical landslide inventory published in the Information System of Mass Movements (SIMMA) by the Colombian Geological Service; DesInventar disaster database (http://www.desinventar.org) and the international disasters database EM-DAT. The results found were isolated cases in the basin that are scarcely documented and correspond to the year 2011; however, no events were reported in the river corridor under analysis. For this reason, no parameter calibration was performed, and the parameters recommended by the literature for the type of soil in the study area were considered. Landslide susceptibility and hazard studies are essential for updating land use plans in the municipalities of the Colombian territory (Aristizábal et al., 2022).

4.2 Landslide propagation (runout) modeling

For landslide propagation assessment, source areas were considered as those areas with FoS <1.1. All the identified high hazard areas are in recent colluvial deposits. It is evident in Figure 5 that as the return period increases, the hazard areas increase; however, for return periods 2.33 and 5 years, the areas of hazards susceptible to landslide with FoS <1.1 are the same; this situation is also present for return periods 25 and 50 years, thus having the same spatial distribution of hazard in term of FoS for the river corridor of analysis, likewise it will have the same propagation footprint in the simulations to determine the runout. These cells were intersected with the soil thickness result to obtain the initial thickness of soil susceptible to landslide as input for the debris flow runout simulations using the RAMMS model.

Figure 6 shows the estimated propagation footprint of the sliding mass emerging from the release sources identified and the maximum deposition heights of the material for the different rainfall scenarios.

The results obtained were conditioned by the spatial resolution of the terrain elevation model. In all scenarios analyzed, the material removed from the release areas reached the Ovejas riverbed. Increases in the extent and heights of material deposition were evident in the river corridor, reaching in all rainfall scenarios a propagation in the riverbed and even in the area where the water intake structure is located. Therefore, it would be convenient to analyze alternatives for mitigating the identified hazards in the design of the project's water intake structures; however, the risk issue is not within the scope of this study.

Once the results of the runout of the debris flow were obtained, the river corridor was evaluated longitudinally to identify those areas of the river channel that should be studied with greater attention and

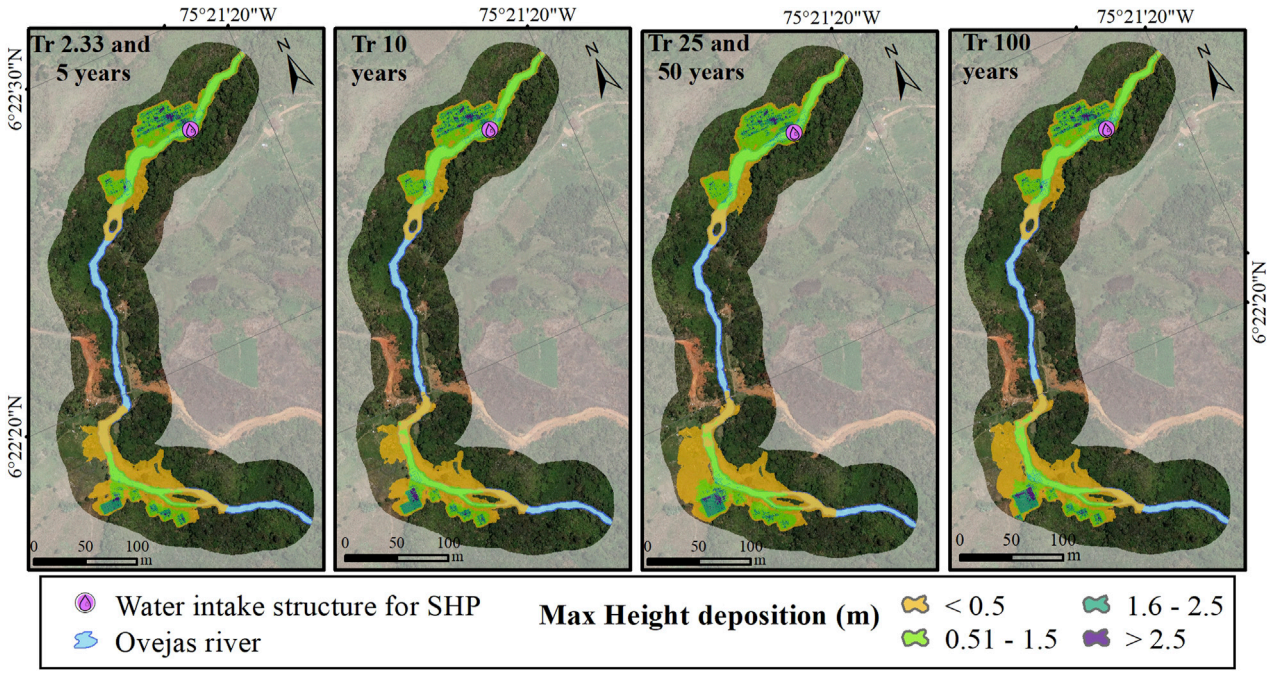


FIGURE 6
Landslide propagation footprint in Ovejas river for each return period.

should require a potential intervention for the protection of the riverbed. The initial point of analysis was defined at the northern end of the project, upstream of the water intake structure for SHP. Figure 7 shows the longitudinal axis of the river in the analysis corridor with the heights of deposition. The color scale corresponds to each return period.

For nearest areas of the sources of material detachment (distances of 100,300,500, and 700 m), there is a high probability that the material will reach the riverbed of the Ovejas River in the return periods analyzed. The section of the river between 500 and 700 m is the most likely to be affected because it is the area with the highest number of sources with FoS <1.1. These critical areas must be managed to mitigate the vulnerability of the intake engineering works in the channel.

To evaluate these results obtained from the modeling of debris flow propagation footprint, the cross section at 600 m in longitudinal profile in the river corridor (just upstream of the SHP project water intake structure) was analyzed to observe the deposition heights that could generate an obstruction of the stream. The results are shown in Figure 8.

Figure 8 shows the heights of material deposition increased simultaneously with the return period analyzed for each scenario

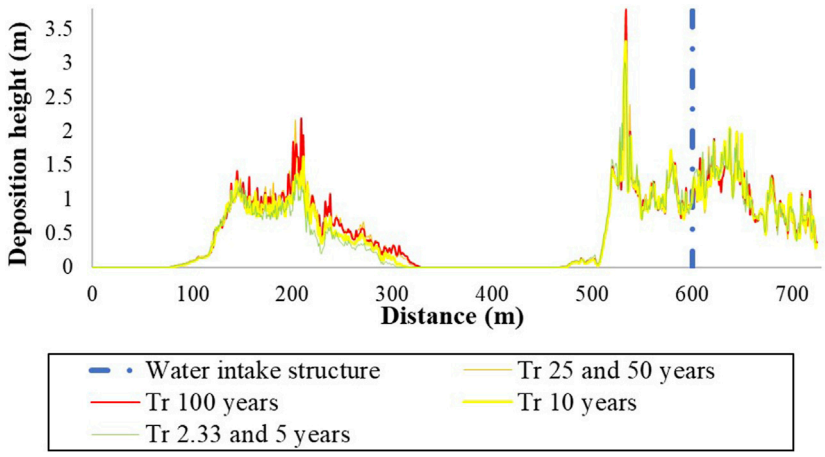


FIGURE 7
Longitudinal profile of deposition height for each return period.

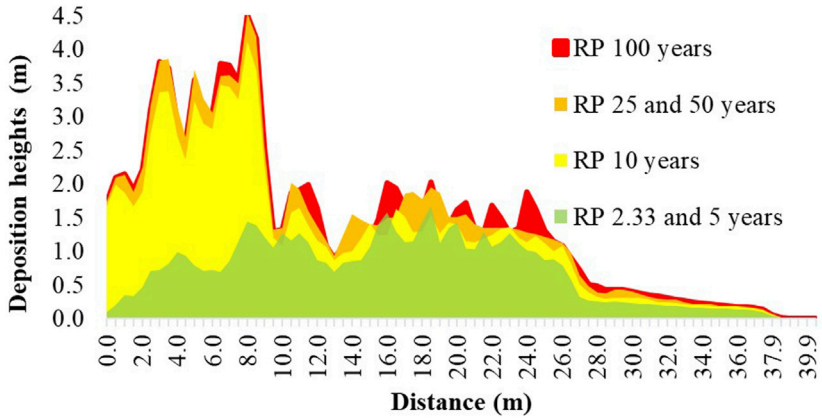


FIGURE 8
Landslide propagation profile in the cross section upstream of the SHP project for each return period.

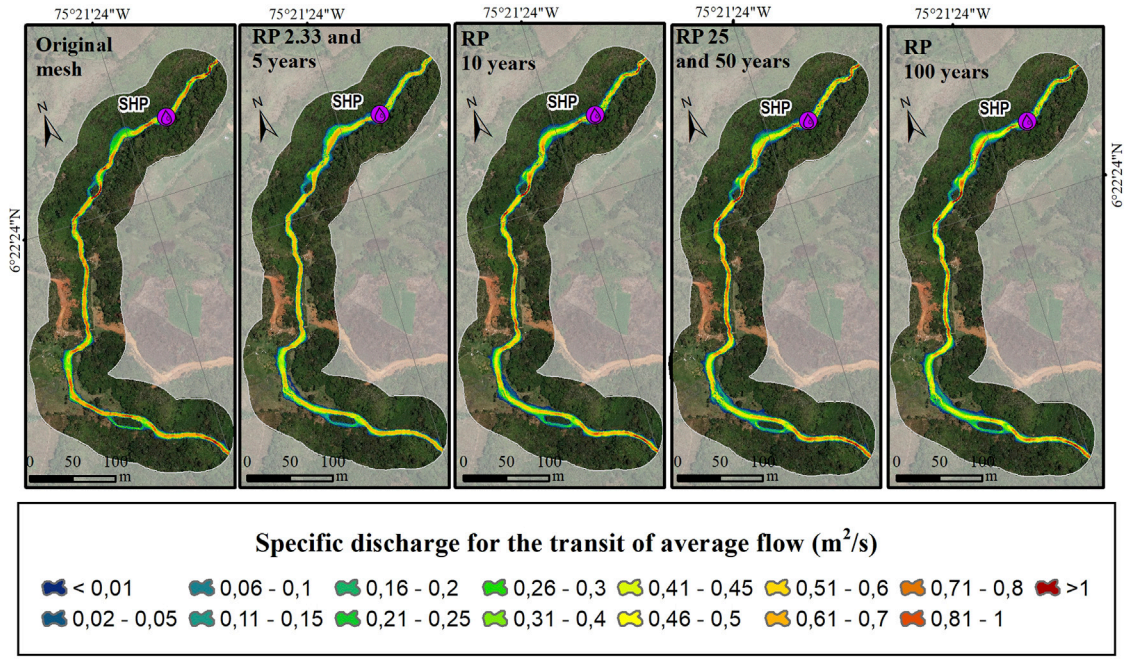


FIGURE 9
Changes in specific discharge for different return period meshes with the transit of average flow.

and the corresponding number of release sources identified for these. These sources were identified as having a high hazard using the SLIDE model. It should be noted that precipitation as a triggering factor is higher as the return period increases but has a lower probability of recurrence.

4.3 Analysis of river flow behavior with landslide dams

This section will show the results of the changes in the specific flow and in the delay times for average flow, being the

TABLE 2 Delay time for average flow.

	Time. Mesh with obstructions (sec)	Additional Time (sec)	Additional Time %
Tr 2.33 and 5 years	780	200	34%
Tr 10 years	810	230	40%
Tr 25 and 50 years	850	270	47%
Tr 100 years	860	280	48%

same flow and different meshes, as well as for each hydrograph, being different hydrographs and different meshes. Delay times refer to the time in which the flow decreases or falls below the average flow or design flow, due to the obstruction of the channel. In a SHP this time may represent an important interval during which the flow for generation falls below the minimum allowed, or in the case of an intake for water supply, this time may imply an interval in which there is a lack of supply.

The first step in the modeling was to obtain the characteristics of the basin and the generation of analysis sub-basins and their drainage network from the GIS module of HEC-HMS, as well as some hydrological parameters. In the meteorological model of HEC-HMS, we applied the hietograms of each of the sub-basins estimated for each return period.

The hydraulic modeling was carried out for different flow and meshes scenarios for the return periods of 2.33, 5, 10, 10, 25, 50, and 100 years, to show the changes in flow height, flow rate and delay time that the flow suffers with the obstruction or landslide dams modeled in RAMMS. The delay time and the decrease in flow are variables that define the hazard due to water shortage in hydroelectric generation.

4.3.1 Stream behavior with channel obstructions for average flow

To identify the delay that can be generated after the obstructions in the channel, the transit of the average flow through the river corridor was modeled with the topography without obstruction (original mesh) and with the topography incorporating the landslide dams generated by the propagation of the debris flow for the different return periods. The specific discharge in [m^2/s] in the grid cell indicates the flow per unit length passing through the cross section of the river. Values of the specific flow indicate a concentration of flow due to the narrowing of the section because of the obstruction. The Changes in specific discharge for different meshes of return period with average flow are shown in Figure 9.

The time delay and the decrease in flow are variables that define the hazard due to water shortage in hydroelectric generation or community supply. The transit time of the average flow in the river corridor was 580 s for the original mesh. Table 2 shows the time of transit for each return period and the additional time required to transit the average flow with respect to the original mesh.

The modeling results shows that the obstructions generated by the chain of landslides and debris flow modified the hydraulic conditions of the river, influencing the flow transit time in the river corridor. On the other hand, the specific discharge was analyzed between the different scenarios in a section upstream of the water intake structure of the project.

For all return periods analyzed there was a decrease in specific flow, with lower values as the return period increased, reaching reductions of more than 50% of the specific flow. This could result in a shortage of water supply, meaning that the project may not be able to generate energy at a given moment due to the absence of the specific flow at the water intake structure site or may leave a community without water supply.

4.3.2 Stream behavior with channel obstructions for hydrographs of different return periods

The channel blockage was caused by a chain of debris flows from rainfall-induced landslides, itself with an associated return period. Under the assumption that the response of the basin will have the same return period, the behavior of the stream in each of the meshes and their corresponding hydrographs according to return the period was modeled. The results are presented in Figures 10–15. Each figure contains in the lower right corner the comparative specific flow rates for each simulation in the upstream section of the intake structure of the project.

The results show that the transit time of the hydrographs in the analyzed river corridor increased for the meshes with landslide dams, its increased time was similar in all return periods, this time increase is equivalent to 11%.The results also showed a decrease in

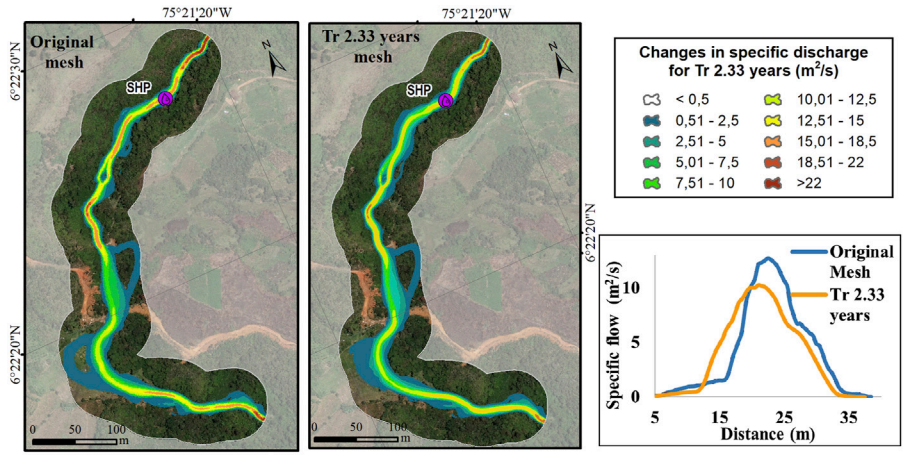


FIGURE 10
Specific discharge for original mesh and for Tr 2.33 years.

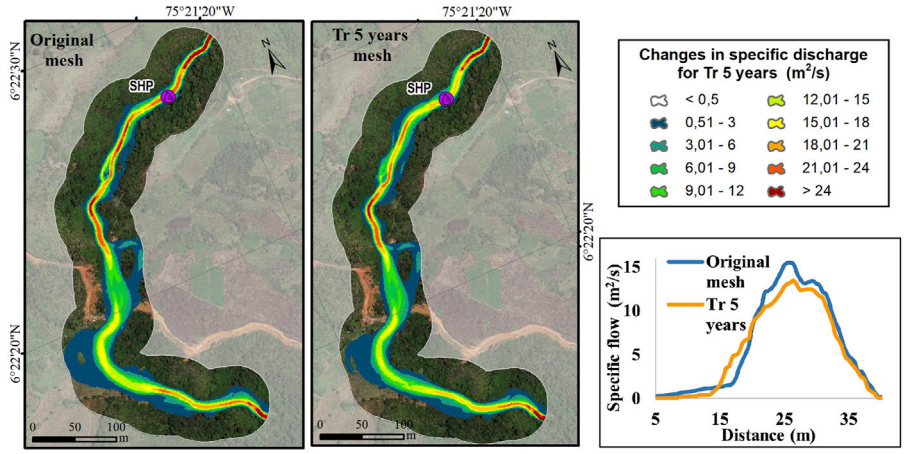


FIGURE 11
Specific discharge for original mesh and for Tr flow of 5 years.

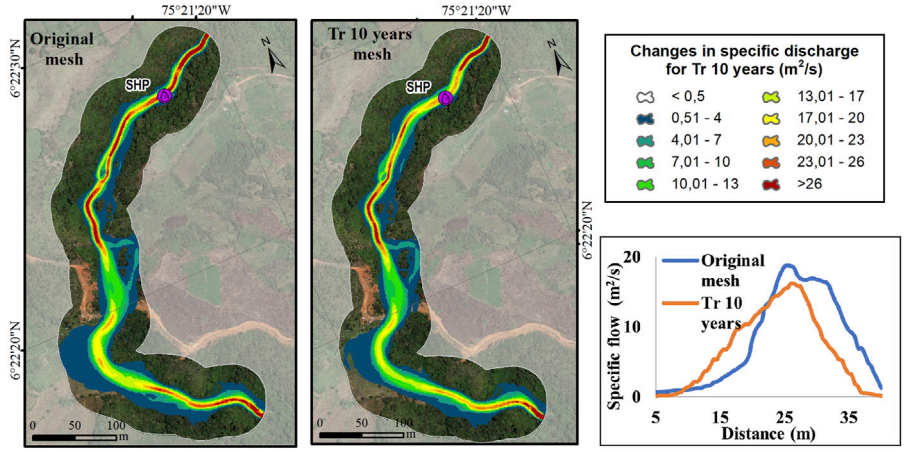


FIGURE 12
Specific discharge for original mesh and for Tr flow of 10 years.

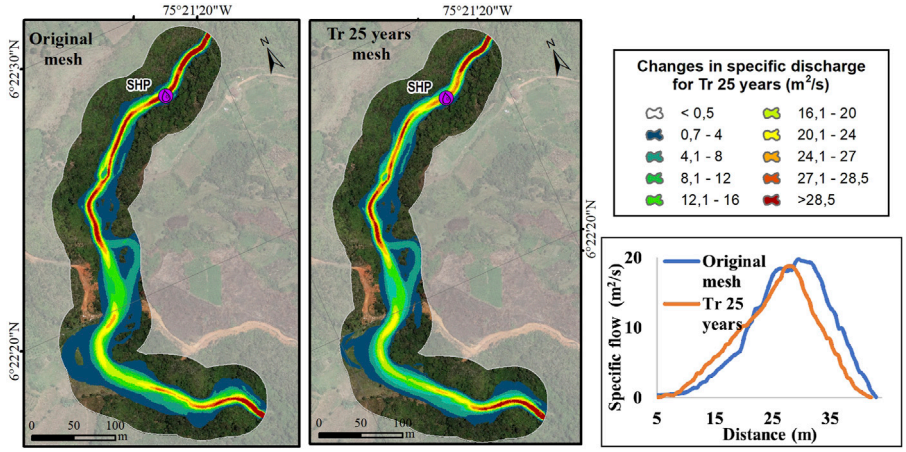


FIGURE 13
Specific discharge for original mesh and for Tr flow of 25 years.

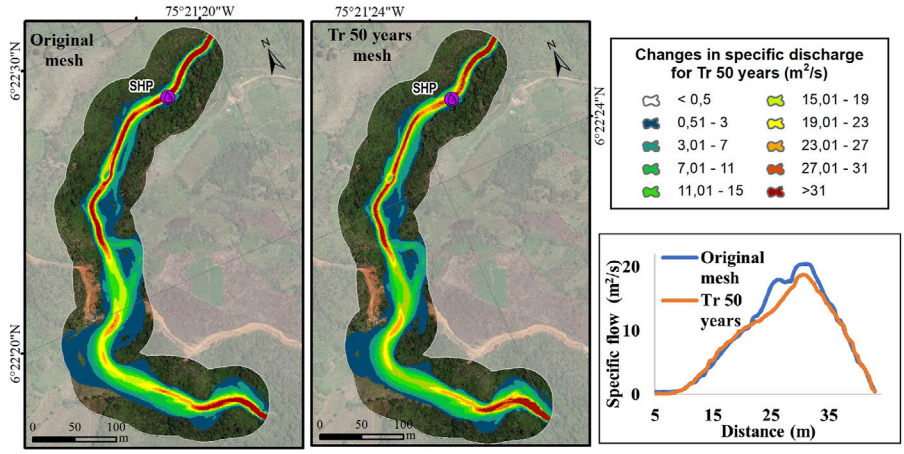


FIGURE 14
Specific discharge for original mesh and for Tr flow of 50 years.

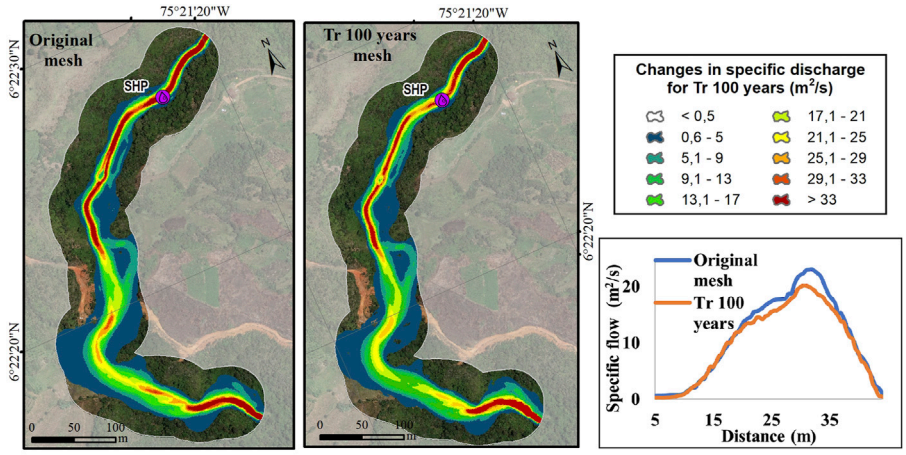


FIGURE 15
Specific discharge for original mesh and for Tr flow of 100 years.

the specific flow in all the scenarios analyzed due to the landslide dams in the river channel, representing up to a 20% decrease in the specific flow.

5 Discussion

The shallow landslides assessment was based on an infinite slope stability model, in which the failure surface is assumed to be flat, in which the uncertainty of the geomechanical parameters of the soil was not considered. The applied methodology is able to simulate rainfall-induced landslides, but it remains difficult to predict the exact time when the slope is prone to instability. In addition, rainfall data can influence the model simulation, especially at the time when landslides occur.

The reliability of the propagation results depends on the determination and calibration of debris flow model parameters that more closely represent the ground conditions and flow characteristics of the material, this process can be performed using two approaches: In the first approach, soil parameters can be obtained from laboratory tests of data collected in the field and are appropriately used and processed for early warning calibration. However, field observations of debris flow behavior and rheology are difficult and still scarce and laboratory experiments are difficult to translate to field situations (Von Boetticher et al., 2017). In the second approach, a retrospective analysis of well-documented past debris flow events is carried out, in which the source zone and propagation zone of the mass movement can be clearly identified to obtain the flow intensity parameters in which calibrated values

of the input parameters can be used to predict the parameters of the potential debris flow in the future, however, in the study area there are no historical records that allow carrying out this analysis, leaving an open path for future research in cases where there are no precedent events for the calibration of the models. The challenge in the application of dynamic models is the estimation or calibration of key model parameters.

The delay or detention times of the river flow upstream of the catchment and the temporary variation of the flow downstream of the obstruction can be critical for the hydroelectric generation of a project, as well as for the supply of a population downstream of the obstruction. This would imply a shortage in the water catchment, which could be reflected in the reduction of energy generation by the hydroelectric plants or the shortage of the aqueduct of a community, which is recommended for further study. As well as the study of possible obstructions that could obstruct and collapse the catchment. This is an issue that should be addressed more rigorously in the future, in the scope of this study it is not possible determine the real impact that this event could have on the generation of energy or the supply of communities.

The evaluation of landslide mitigation measures is very complex and must consider more aspects. In this article, we do not intend to make an exhaustive review of the landslide dam that may be generated. To offer a deeper discussion, we only focus on the chain as such that can be triggered and its possible effects, especially to the water supply shortage of the SHP under study or in other possible small hydroelectric powers and communal aqueducts, which are the main users of rural basis in this type of mountainous areas, which can be a good attempt to encourage the scientific communities as potential future research step.

6 Conclusion

In this work, a comprehensive assessment of landslide dam by the occurrence of rainfall induced landslide debris flow hazard chain using deterministic method was presented. SLIDE model was implemented for develop a stability analysis in terms of FoS through multiple rainfall return periods, identifying high hazard zones that may detach sliding material and generate obstructions in the river channel. Medium and high hazards were observed in the areas of geological units of recent colluvial deposits that are coincident with high slopes. Hazard increases as the return period increases. It is important to note that although slopes are a factor that probably have a high degree of influence, they do not control the hazard zoning and are affected by soil parameters, including the thickness susceptible to landslide. The lower FoS values are mainly distributed towards areas with thicker landslide-susceptible soils, where more intense precipitation events progressively affect soil saturation and possible collapse, which is consistent with the model and its intention to simulate natural events. For all the scenarios analyzed, there is a predominance of low to exceptionally low hazard levels in the Antioquia Batholith.

RAMMS-DF was used to simulate debris flow. The sources identified as high hazard zones propagated the material until it reached the riverbed, generating partial obstructions in the analysis section located near the water intake structure for the Small Hydropower Plant. However, in not one of the scenarios analyzed was there a total obstruction of the river cross section according to the morphology of the channel and the heights deposited on the riverbed.

The hazard of obstruction of water sources due to the occurrence of landslides dams was evaluated in terms of the retention of the flow. It was found that when evaluating the hydraulic behavior of the river channel against landslide dams generated by the occurrence of rainfall-induced landslide and debris flow hazard chain for the different return periods, the obstructions were sufficient to change the hydraulic regime of the study river corridor showing longer delay times in the flow transit, in addition to the decrease in the specific flow. For average flow conditions with obstructions in the riverbed, the specific flow rate showed reductions of more than 50%. For hydrographs of different return periods, the results show a decrease in specific flow in all scenarios analyzed, up to 20%. Additionally, for these hydrographs, the results show that the transit time increased by up to 11%.

These results have potential application in the management of slope disaster prevention and the assessment of the hazard chain of rainfall-induced shallow landslides and the consequent generation of debris flows. (Han et al., 2021).

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

LO-G developed the theory and performed the computations. All authors contributed to the study conception and design BB and JV verified the analytical methods. All authors contributed to the article and approved the submitted version.

Acknowledgments

The proposed methodology in this study is framed within the guidelines and products of the research program “Vulnerability, resilience and risk of communities and supplying basins affected by landslides and avalanches,” code 1118-852-71251, project “Functions for vulnerability assessment due to water shortages by landslides and avalanches: micro-basins of southwest Antioquia”, contract 80740-492-2020 held between Fiduprevisora and the Universidad de Medellín, with resources from the National Financing Fund for science, technology and innovation, “Francisco José de Caldas”. We also thank Integral S.A. for the information provided of the Guamito Hydroelectric Project Feasibility Study developed in 2019. PALSAR DEM data was downloaded from the Alaska Satellite Facility (ASF).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1157881/full#supplementary-material

References

Abraham, M. T., Satyam, N., Pradhan, B., and Tian, H. (2022). Debris flow simulation 2D (DFS 2D): Numerical modelling of debris flows and calibration of friction parameters. *J. Rock Mech. Geotechnical Eng.* 14, 1747–1760. doi:10.1016/J.JRMGE.2022.01.004

AMVAÁrea Metropolitana del Valle de AburráCorporaciones Autónomas RegionalesCorporación Autónoma Regional del Centro de AntioquiaCorporación Autónoma Regional de las Cuencas de los Ríos Negro y Nare (2015). Plan de Ordenación y Manejo de la Cuenca Hidrográfica del Río Aburrá. Available at: https://www.corantioquia.gov.co/planes-de-ordenacion-y-manejo-de-la-cuenca-hidrografica-pomca/ (Accessed May 7, 2023).

Arghya, A. B., Hawlader, B., and Guthrie, R. H. (2022). "A comparison of two runout programs for debris flow assessment at the Solalex-Anzeindaz region of Switzerland," in *Géorisques - VIII - Geohazards* (Quebec, Canada). Available at: https://www.stantec.com/en/ideas/a-comparison-of-two-runout-programs-for-debris-flow-assessment-solalex-anzeindaz-region-switzerland (Accessed May 1, 2023).

Aristizábal, E., Garcia, E. F., Marin, R. J., Gómez, F., Guzmán-Martínez, J., Aristizábal, E., et al. (2022). *Rainfall-intensity effect on landslide hazard assessment due to climate change in north-western Colombian Andes*. Medellín, Colombia: Revista Facultad de Ingeniería Universidad de Antioquia, 51–66. doi:10.17533/UDEA.REDIN.20201215

Aristizábal, E., Martínez-Carvajal, H., and García-Aristizábal, E. (2017). Modelling shallow landslides triggered by rainfall in tropical and mountainous basins. *Adv. Cult. Living Landslides* 207, 207–212. doi:10.1007/978-3-319-53485-5_23

Armaş, I., Gheorghe, M., and Silvaş, G. C. (2021). Shallow landslides physically based susceptibility assessment improvement using InSAR. Case study: Carpathian and subcarpathian prahova valley, Romania. *Remote Sens.* 13, 2385. doi:10.3390/RS13122385

ASF DAAC (2015). ALOS PALSAR_Radiometric_Terrain_Corrected_high_res. Available at: https://asf.alaska.edu/data-sets/derived-data-sets/alos-palsar-rtc/alos-palsar-radiometric-terrain-correction/ (Accessed November 11, 2021).doi:10.5067/Z97HFCNKR6VA

Baggio, T., Mergili, M., and D'Agostino, V. (2021). Advances in the simulation of debris flow erosion: The case study of the Rio Gere (Italy) event of the 4th August 2017. *Geomorphology* 381, 107664. doi:10.1016/J.GEOMORPH.2021.107664

Bladé, E., Cea, L., Corestein, G., Escolano, E., Puertas, J., Vázquez-Cendón, E., et al. (2014). Iber — river modelling simulation tool. *Rev. Int. Métodos Numéricos Cálculo Diseño Ing.* 30, 1–10. doi:10.1016/J.RIMNI.2012.07.004

Bomers, A. (2020). Hydraulic modelling approaches to decrease uncertainty in flood frequency relations. PhD Thesis. (University of Twente: Enschede, Netherlands) doi:10.3990/1.9789036549288

Calcaterra, D., Martire, D. Di, Guerriero, L., Tomás, R., Košová, V., Molokáč, M., et al. (2022). Avalanche hazard modelling within the kráľova hoľa area in the low tatra mountains in Slovakia. *Land* 11, 766. doi:10.3390/LAND11060766

Chae, B. G., Park, H. J., Catani, F., Simoni, A., and Berti, M. (2017). Landslide prediction, monitoring and early warning: A concise review of state-of-the-art. *Geosciences J.* 21 (6), 1033–1070. doi:10.1007/S12303-017-0034-4

Chen, W., Tsai, F., Kainz, W., Thomas, J., Gupta, M., Srivastava, P. K., et al. (2023). Assessment of a dynamic physically based slope stability model to evaluate timing and distribution of rainfall-induced shallow landslides. *ISPRS Int. J. Geo-Information* 12, 105. doi:10.3390/IJGI12030105

Chen, Y., Zhao, L., Wang, Y., Jiang, Q., and Qi, D. (2019). Precipitation data and their uncertainty as input for rainfall-induced shallow landslide models. *Front. Earth Sci.* 13, 695–704. doi:10.1007/s11707-019-0791-7

Chen, Z., He, S., Shen, W., and Wang, D. (2022). Effects of defense-structure system for bridge piers on two-phase debris flow wakes. *Acta Geotech.* 17, 1645–1665. doi:10.1007/s11440-021-01296-5

Chikalamo, E. E. (2018). Comparing modelling approaches for landside early warning: A case study of bogowonto catchment, central java, Indonesia, Thesis. Enschede, The Netherlands: ITC Faculty Geo-Information Science and Earth Observation.

Christen, M., Kowalski, J., and Bartelt, P. (2010). Ramms: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Reg. Sci. Technol.* 63, 1–14. doi:10.1016/J.COLDREGIONS.2010.04.005

Cui, P., and Guo, J. (2021). Evolution models, risk prevention and control countermeasures of the valley disaster chain. *Gongcheng Kexue Yu Jishu/Advanced Eng. Sci.* 53, 5–18. doi:10.15961/J.JSUESE.202100285

Dash, R. K., Samanta, M., and Kanungo, D. P. (2023). "Debris flow hazard in India: Current status, research trends, and emerging challenges," in *Landslides: Detection, Prediction and Monitoring* Editors P. Thambidurai and T.N. Singh (Manhattan, NY, USA: Springer, Cham), 211–231. doi:10.1007/978-3-031-23859-8_10

Dhanai, P., Singh, V. P., and Soni, P. (2022). Rainfall triggered slope instability analysis with changing climate. *Indian Geotechnical J.* 52, 477–492. doi:10.1007/s40098-021-00581-0

Díaz-Salas, A. M., Guevara-Pérez, E., and Vidal-Moreno, J. D. (2021). Modelamiento numérico de un flujo de escombros asociado a una rotura de presa en la subcuenca Quillcay, Áncash, Perú. *Rev. Ing. UC* 28, 35–46. doi:10.54139/revinguc.v28i1.4

Do, H. M., Long, K., Zi, Y., Guo, Z., Yin, K. L., and Guo, Z. Z. (2020). A comparative study on the integrative ability of the analytical hierarchy process, weights of evidence and logistic regression methods with the Flow-R model for landslide susceptibility assessment. *Geomatics, Nat. Hazards Risk* 11 (1), 2449–2485. doi:10.1080/19475705.2020.1846086

Dysarz, T. (2018). Development of RiverBox—an ArcGIS toolbox for river bathymetry reconstruction. *Water* 10, 1266. doi:10.3390/W10091266

Fernandes Azevedo, G., Montoya Botero, E., Martínez, H. E., García, E., and Newton, M. S. (2015). "Estimativa da profundidade do solo pelo uso de técnicas de geoprocessamento, estudo de caso: Setor Pajarito, Colômbia," in XVII Brazilian symposium on remote sensing. João pessoa, Brazil April 2015.

Franco-Ramos, O., Ballesteros-Cánovas, J. A., Figueroa-García, J. E., Vázquez-Selem, L., Stoffel, M., and Caballero, L. (2020). Modelling the 2012 lahar in a sector of jamapa gorge (pico de Orizaba volcano, Mexico) using RAMMS and tree-ring evidence. *Water* 12, 333. doi:10.3390/W12020333

Fustos-Toribio, I. J., Morales-Vargas, B., Somos-Valenzuela, M., Moreno-Yaeger, P., Muñoz-Ramirez, R., Rodriguez Araneda, I., et al. (2021). Debris flow event on Osorno volcano, Chile, during summer 2017: New interpretations for chain processes in the southern Andes. *Nat. Hazards Earth Syst. Sci.* 21, 3015–3029. doi:10.5194/NHESS-21-3015-2021

Gan, J., and Zhang, Y. X. (2019). Numerical simulation of debris flow runout using ramms: A case study of luzhuang gully in China. *Comput. Model Eng. Sci.* 121, 981–1009. doi:10.32604/cmes.2019.07337

García, J. V. G., Panta, J. E. R., Reynoso, D. S. F., Ayala, C. R., Hidalgo, R. R., García, F. G. G., et al. (2022). Modelación hidráulica en Iber para prevención de inundaciones en la cuenca Tesechoacán. *Rev Mex Cienc* 13, 159–181. doi:10.29298/RMCF.V13I71.1238

García-Alén, G., González-Cao, J., Fernández-Nóvoa, D., Gómez-Gesteira, M., Cea, L., and Puertas, J. (2022). Analysis of two sources of variability of basin outflow hydrographs computed with the 2D shallow water model Iber: Digital Terrain Model and unstructured mesh size. *J. Hydrol. (Amst)* 612, 128182. doi:10.1016/J.JHYDROL.2022.128182

García-Delgado, H., Villamizar-Escalante, N., Bermúdez, M. A., Bernet, M., and Velandia, F. (2021). Climate or tectonics? What controls the spatial-temporal variations in erosion rates across the eastern cordillera of Colombia? *Glob. Planet Change* 203, 103541. doi:10.1016/J.GLOPLACHA.2021.103541

González-Cao, J., García-Feal, O., Fernández-Nóvoa, D., Domínguez-Alonso, J. M., and Gómez-Gesteira, M. (2019). Towards an automatic early warning system of flood

hazards based on precipitation forecast: The case of the miño river (NW Spain). *Nat. Hazards Earth Syst. Sci.* 19, 2583–2595. doi:10.5194/NHESS-19-2583-2019

Graf, C., and McArdell, B. (2008). "Simulation of debris flow runout before and after construction of mitigation measures: an example from the Swiss Alps," in International Conference *Debris Flows: Disasters, Risk, Forecast, Protection* (Pyatigorsk, Russia), 233–236. Available at: https://elibrary.ru/item.asp?id=37236604 (Accessed May 7, 2023).

Guo, J., Cui, Y., Xu, W., Yin, Y., Li, Y., and Jin, W. (2022a). Numerical investigation of the landslide-debris flow transformation process considering topographic and entrainment effects: A case study. *Landslides* 19, 773–788. doi:10.1007/s10346-021-01791-6

Guo, Z., Torra, O., Hürlimann, M., Abancó, C., and Medina, V. (2022b). Fslam: A qgis plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environ. Model. Softw.* 150, 105354. doi:10.1016/J.ENVSOFT.2022.105354

Hafnaoui, M. A., and Debabeche, M. (2021). Numerical modeling of the hydraulic jump location using 2D Iber software. *Model Earth Syst. Environ.* 7, 1939–1946. doi:10.1007/s40808-020-00942-3

Hamdan, A. N. A., Almuktar, S., and Scholz, M. (2021). Rainfall-runoff modeling using the HEC-HMS model for the Al-adhaim river catchment, northern Iraq. *Hydrology* 8, 58. doi:10.3390/HYDROLOGY8020058

Hidalgo, C. A., and Vega, J. A. (2021). Probabilistic landslide risk assessment in water supply basins: La liboriana River Basin (salgar-Colombia). *Nat. Hazards* 109, 273–301. doi:10.1007/s11069-021-04836-0

Hoyos, H., and Botero, B. A. (2019). Vulnerability assessment with scarce information for a quantitative flood risk model. Case study monteria-Colombia. *IOP Conf. Ser. Mater Sci. Eng.* 471, 102005. doi:10.1088/1757-899X/471/10/102005

Hu, H., Zhou, G. G. D., Song, D., Cui, K. F. E., Huang, Y., Choi, C. E., et al. (2020). Effect of slit size on the impact load against debris-flow mitigation dams. *Eng. Geol.* 274, 105764. doi:10.1016/J.ENGGEO.2020.105764

Hungr, O., Leroueil, S., and Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides* 11, 167–194. doi:10.1007/s10346-013-0436-y

Hürlimann, M., Guo, Z., Puig-Polo, C., and Medina, V. (2022). Impacts of future climate and land cover changes on landslide susceptibility: Regional scale modelling in the val d'Aran region (pyrenees, Spain). *Landslides* 19, 99–118. doi:10.1007/s10346-021-01775-6

Kun-Ting, C., Chia-Hsing, L., Xiao-Qing, C., Gui-Sheng, H., Xiao-Jun, G., Chjeng-Lun, S., et al. (2018). An assessment method for debris flow dam formation in taiwan. *Earth Sci. Res. J.* 22, 37–43. doi:10.15446/ESRJ.V22N1.62389

Li, Y., Chen, J., Zhou, F., Song, S., Zhang, Y., Gu, F., et al. (2020a). Identification of ancient river-blocking events and analysis of the mechanisms for the formation of landslide dams in the Suwalong section of the upper Jinsha River, SE Tibetan Plateau. *Geomorphology* 368, 107351. doi:10.1016/J.GEOMORPH.2020.107351

Li, Y., Chen, J., Zhou, F., Song, S., Zhang, Y., Gu, F., et al. (2020b). Identification of ancient river-blocking events and analysis of the mechanisms for the formation of landslide dams in the Suwalong section of the upper Jinsha River, SE Tibetan Plateau. *Geomorphology* 368, 107351. doi:10.1016/J.GEOMORPH.2020.107351

Liao, Z., Hong, Y., Adler, R. F., and Bach, D. (2011). "A physically based SLIDE model for landslide hazard assessments using remotely sensed data sets," in *Geomechanics and geotechnics* (Boca raton, FL, USA: CRC Press), 807–813. doi:10.1201/B10528-127

Liao, Z., Hong, Y., Wang, J., Fukuoka, H., Sassa, K., Karnawati, D., et al. (2010). Prototyping an experimental early warning system for rainfall-induced landslides in Indonesia using satellite remote sensing and geospatial datasets. *Landslides* 7, 317–324. doi:10.1007/s10346-010-0219-7

Liu, T., Wang, Y., Yu, H., and Chen, Y. (2022). Using statistical functions and hydro-hydraulic models to develop human vulnerability curves for flash floods: The flash flood of the Taitou catchment (China) in 2016. *Int. J. Disaster Risk Reduct.* 73, 102876. doi:10.1016/J.IJDRR.2022.102876

Liu, X., Zhao, C., Zhang, Q., Lu, Z., Li, Z., Yang, C., et al. (2021). Integration of Sentinel-1 and ALOS/PALSAR-2 SAR datasets for mapping active landslides along the Jinsha River corridor, China. *Eng. Geol.* 284, 106033. doi:10.1016/J.ENGGEO.2021.106033

Marin, R. J., Velásquez, M. F., and Sánchez, O. (2021). Applicability and performance of deterministic and probabilistic physically based landslide modeling in a data-scarce environment of the Colombian Andes. *J. South Am. Earth Sci.* 108, 103175. doi:10.1016/J.JSAMES.2021.103175

Medina, V., Hürlimann, M., Guo, Z., Lloret, A., and Vaunat, J. (2021). Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *Catena (Amst)* 201, 105213. doi:10.1016/J.CATENA.2021.105213

Mikoš, M., and Bezak, N. (2021). Debris flow modelling using RAMMS model in the alpine environment with focus on the model parameters and main characteristics. *Front. Earth Sci. (Lausanne)* 8, 732. doi:10.3389/feart.2020.605061

Montoya Botero, E. (2018). Metodologia para aplicação de redes neurais artificiais para sistemas de alerta de escorregamentos deflagrados por chuvas em regiões montanhosas. Available at: https://repositorio.unb.br/handle/10482/33056 (Accessed May 7, 2023).

Montrasio, L., Valentino, R., Losi, G. L., Corina, A., Rossi, L., and Rudari, R. (2013). Space-time hazard assessment of rainfall-induced shallow landslides. *Landslide Sci. Pract. Glob. Environ. Change* 4, 283–293. doi:10.1007/978-3-642-31337-0_37

Moreira Melo, C., Kobiyama, M., Paulo Michel, G., Madruga de Brito, M., Lu, Z., and Miguel Ferreira, T. (2021). The relevance of geotechnical-unit characterization for landslide-susceptibility mapping with SHALSTAB. *GeoHazards* 2, 383–397. doi:10.3390/GEOHAZARDS2040021

Náquira Bazán, M. V. (2009). Susceptibilidad de Remociones en Masa en las Costas de Fiordos Cercanos a Hornopirén, X Región. Available at: https://repositorio.uchile.cl/handle/2250/103473 (Accessed May 7, 2023).

Nian, T., Li, D., Liang, Q., Wu, H., and Guo, X. (2021). Multi-phase flow simulation of landslide dam formation process based on extended coupled DEM-CFD method. *Comput. Geotech.* 140, 104438. doi:10.1016/J.COMPGEO.2021.104438

Oliveira, S. C., Zêzere, J. L., Lajas, S., and Melo, R. (2017). Combination of statistical and physically based methods to assess shallow slide susceptibility at the basin scale. *Nat. Hazards Earth Syst. Sci.* 17, 1091–1109. doi:10.5194/NHESS-17-1091-2017

Pérez-Montiel, J. I., Cardenas-Mercado, L., and Nardini, A. G. C. (2022). Flood modeling in a coastal town in northern Colombia: Comparing MODCEL vs. IBER. *Water* 14, 3866. doi:10.3390/W14233866

Pulgarín Dávila, E. G. (2009). Fórmulas regionales para la estimación de curvas intensidad-frecuencia-duración basadas en las propiedades de escala de la lluvia (región andina colombiana). Available at: https://repositorio.unal.edu.co/bitstream/handle/unal/70277/98671272.2009_1.pdf?sequence=3 (Accessed May 1, 2023).

Pulgarín, E., and Poveda, G. (2008). "Estimación de curvas IDF basadas en las propiedades de escala de la lluvia," in Proceedings of the XVIII Seminario Nacional de Hidráulica e Hidrología. Bogotá, Colombia May 2008.

Qiang, X., Dalei, P., Chaoyang, H., Xing, Q., Kuanyao, Z., Dehao, X., et al. (2020). Theory and method of monitoring and early warning for sudden loess landslide—A case study at heifangtai terrace. *J. Eng. Geol.* 28 (1), 111–121. doi:10.13544/J.CNKI.JEG.2019-038

Rana, H., and Babu, G. L. S. (2022). Regional back analysis of landslide events using TRIGRS model and rainfall threshold: An approach to estimate landslide hazard for kodagu, India. *Bull. Eng. Geol. Environ.* 81, 160–216. doi:10.1007/s10064-022-02660-9

Reddy, B. S. N., and Pramada, S. K. (2022). A hybrid artificial intelligence and semi-distributed model for runoff prediction. *Water Supply* 22, 6181–6194. doi:10.2166/WS.2022.239

Roldán, F., Salazar, I., González, G., Roldán, W., and Toro, N. (2022). Flow-type landslides analysis in arid zones: Application in La chimba basin in antofagasta, atacama desert (Chile). *Water* 14, 2225. doi:10.3390/W14142225

Ruiz-Villanueva, V., Gamberini, C., Bladé, E., Stoffel, M., and Bertoldi, W. (2020). Numerical modeling of instream wood transport, deposition, and accumulation in braided morphologies under unsteady conditions: Sensitivity and high-resolution quantitative model validation. *Water Resour. Res.* 56, e2019WR026221. doi:10.1029/2019WR026221

Sahu, S. A. L. I. L., Pyasi, S. K., and Galkate, R. V. (2020). A review on the HEC-HMS rainfall–runoff simulation model. *Int. J. Agric. Sci. Res.* 10 (4), 183–190.

Thomas, J., Gupta, M., Srivastava, P. K., and Petropoulos, G. P. (2023). Assessment of a dynamic physically based slope stability model to evaluate timing and distribution of rainfall-induced shallow landslides. *ISPRS Int. J. Geoinf* 12, 105. doi:10.3390/IJGI12030105

Trujillo-Vela, M. G., Ramos-Cañón, A. M., Escobar-Vargas, J. A., and Galindo-Torres, S. A. (2022). An overview of debris-flow mathematical modelling. *Earth Sci. Rev.* 232, 104135. doi:10.1016/J.EARSCIREV.2022.104135

Tyagi, A., Kamal Tiwari, R., and James, N. (2022). A review on spatial, temporal and magnitude prediction of landslide hazard. *J. Asian Earth Sci. X* 7, 100099. doi:10.1016/J.JAESX.2022.100099

UNISDRCRED (2017). *Economic losses, poverty and disasters 1998-2017*. (Geneva, Switzerland: UNDRR). doi:10.13140/RG.2.2.35610.08643

Vega, J. A., and Hidalgo, C. A. (2021). "Methodology for landslides assessment causing river channel obstructions and the consequent water shortage in rural communities," in *ICL contribution to landslide disaster risk reduction* (Manhattan, NY, USA: Springer, Cham). doi:10.1007/978-3-030-60227-7_32

Vega, J. A., and Hidalgo, C. A. (2016). Quantitative risk assessment of landslides triggered by earthquakes and rainfall based on direct costs of urban buildings. *Geomorphology* 273, 217–235. doi:10.1016/J.GEOMORPH.2016.07.032

Von Boetticher, A., Turowski, J. M., Mcardell, B. W., Rickenmann, D., Hürlimann, M., Scheidl, C., et al. (2017). DebrisInterMixing-2.3: A finite volume solver for three-dimensional debris-flow simulations with two calibration parameters-Part 2: Model validation with experiments. *Geosci. Model Dev.* 10, 3963–3978. doi:10.5194/gmd-10-3963-2017

Walczak, N., Walczak, Z., and Nieć, J. (2021). Influence of debris on water intake gratings in small hydroelectric plants: An experimental study on hydraulic parameters. *Energies* 14, 3248. doi:10.3390/EN14113248

WSL Institute for SnowAvalanche Research SLF (2022). RAMMS::DEBRISFLOW User Manual A numerical model for debris flows in research and practice User Manual v1.8.0 Debris Flow RAMMS rapid mass movement simulation Title picture: Debris Flow at Illgraben, Available at: https://ramms.slf.ch/ramms/downloads/RAMMS_DBF_Manual.pdf.

Zeng, P., Wang, S., Sun, X., Fan, X., Li, T., Wang, D., et al. (2022). Probabilistic hazard assessment of landslide-induced river damming. *Eng. Geol.* 304, 106678. doi:10.1016/J.ENGGEO.2022.106678

Zhang, J., Qiu, H., Tang, B., Yang, D., Liu, Y., Liu, Z., et al. (2022a). Accelerating effect of vegetation on the instability of rainfall-induced shallow landslides. *Remote Sens. (Basel)* 14, 5743. doi:10.3390/RS14225743

Zhang, X., Li, Lei, and Xu, C. (2022b). Large-scale landslide inventory and their mobility in lvliang city, shanxi province, China. *Nat. Hazards Res.* 2, 111–120. doi:10.1016/J.NHRES.2022.05.002

Zhang, Y., Xu, X., Li, Z., Yi, R., Xu, C., and Luo, W. (2022c). Modelling soil thickness using environmental attributes in karst watersheds. *Catena (Amst)* 212, 106053. doi:10.1016/J.CATENA.2022.106053

Zhou, W., Qiu, H., Wang, L., Pei, Y., Tang, B., Ma, S., et al. (2022). Combining rainfall-induced shallow landslides and subsequent debris flows for hazard chain prediction. *Catena (Amst)* 213, 106199. doi:10.1016/J.CATENA.2022.106199

Zhu, L., He, S., Qin, H., He, W., Zhang, H., Zhang, Y., et al. (2021). Analyzing the multi-hazard chain induced by a debris flow in Xiaojinchuan River, Sichuan, China. *Eng. Geol.* 293, 106280. doi:10.1016/J.ENGGEO.2021.106280

Zimmermann, F., McArdell, B. W., Rickli, C., and Scheidl, C. (2020). 2D runout modelling of hillslope debris flows, based on well-documented events in Switzerland. *Geosciences* 10, 70. doi:10.3390/GEOSCIENCES10020070

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 21 July 2023
DOI 10.3389/feart.2023.1105610

OPEN ACCESS

EDITED BY
Zefa Yang,
Central South University, China

REVIEWED BY
Chen Shuai,
Central South University, China
Jianxiu Wang,
Tongji University, China

*CORRESPONDENCE
Mahmood Ahmad,
ahmadm@uetpeshawar.edu.pk

RECEIVED 22 November 2022
ACCEPTED 13 July 2023
PUBLISHED 21 July 2023

CITATION
Ahmad M, Alsulami BT, Hakamy A, Majdi A, Alqurashi M, Sabri Sabri MM, Al-Mansob RA and Bin Ibrahim MR (2023), The performance comparison of the decision tree models on the prediction of seismic gravelly soil liquefaction potential based on dynamic penetration test. *Front. Earth Sci.* 11:1105610. doi: 10.3389/feart.2023.1105610

The performance comparison of the decision tree models on the prediction of seismic gravelly soil liquefaction potential based on dynamic penetration test

Mahmood Ahmad[1,2]*, Badr T. Alsulami[3], Ahmad Hakamy[3], Ali Majdi[4], Muwaffaq Alqurashi[5], Mohanad Muayad Sabri Sabri[6], Ramez A. Al-Mansob[1] and Mohd Rasdan Bin Ibrahim[7]

[1]Department of Civil Engineering, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak, Malaysia, [2]Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu, Pakistan, [3]Department of Civil Engineering, College of Engineering and Islamic Architecture, Umm Al-Qura University, Makkah, Saudi Arabia, [4]Department of Building and Construction Techniques Engineering, Al-Mustaqbal University College, Al-Hilla, Iraq, [5]Department of Civil Engineering, College of Engineering, Taif University, Taif, Saudi Arabia, [6]Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia, [7]Department of Civil Engineering, Center for Transportation Research, Engineering Faculty, Universiti Malaya, Kuala Lumpur, Malaysia

Seismic liquefaction has been reported in sandy soils as well as gravelly soils. Despite sandy soils, a comprehensive case history record is still lacking for developing empirical, semi-empirical, and soft computing models to predict this phenomenon in gravelly soils. This work compiles documentation from 234 case histories of gravelly soil liquefaction from across the world to generate a database, which will then be used to develop seismic gravelly soil liquefaction potential models. The performance measures, namely, accuracy, precision, recall, F-score, and area under the receiver operating characteristic curve, were used to evaluate the training and testing tree-based models' performance and highlight the capability of the logistic model tree over reduced error pruning tree, random tree and random forest models. The findings of this research can provide theoretical support for researchers in selecting appropriate tree-based models and improving the predictive performance of seismic gravelly soil liquefaction potential.

KEYWORDS

gravelly soil, liquefaction, reduced error pruning tree, random forest, dynamic penetration test, logistic model tree, random tree

1 Introduction

Liquefaction occurs when a saturated soil loses its strength due to a rise in pore water pressure caused by dynamic loads. It is a condition in which earthquake shaking or other rapid loading weakens the stiffness and strength of a soil. When soil liquefies, it loses its strength and the ability of a soil deposit to sustain the structure above it. This phenomenon causes destructions to environment, structures and human life. Geotechnical engineers must examine soil liquefaction characteristics as part of their profession, in the design stage of civil engineering project (Ghani et al., 2021). Most prior studies concentrated on developing liquefaction evaluation models for sandy or silty soils while ignoring the potential of

liquefaction in gravelly soils. Because gravelly soils are very permeable and contain big particles, extra pore water pressure cannot build up quickly during earthquake loading. Many studies of historical liquefaction-induced risks, however, have revealed that loose to medium density gravelly soil can also liquefy after major earthquakes (Youd et al., 1985; Yegian et al., 1994; Sirovich, 1996; Hatanaka et al., 1997; Lin et al., 2004; Cao et al., 2011).

Recent research on the mechanism and behavior of gravelly soil liquefaction has revealed that the triggering conditions for gravelly soil liquefaction differ from those for sandy or silty soil (Wang and Wang, 2017; Chen et al., 2018; Hu, 2021a). For example, a review of the majority of the soil profile reveals a thick non-liquefiable sandy gravel layer with high penetration resistance (potentially indicating a dense soil deposit) that may have acted as a less-permeable capping layer, contributing to the development of high excess pore water pressures leading to liquefaction while also preventing sand ejecta from reaching the ground surface and liquefaction manifestation to be observed. Because of these differences, gravel soil liquefies differently than sandy or silty soil. Researchers and engineers are attempting to assess the liquefaction potential in this type of soil owing to the widespread occurrence of seismic gravelly soil liquefaction during major earthquakes around the world. However, there is a scarcity of case histories of gravelly soil liquefaction created to develop capable models (Yegian et al., 1994; Hatanaka et al., 1997).

Andrus and Stokoe (2000) proposed the first simplified approach based on V_s and cyclic resistance ratio (CRR) to evaluate the liquefaction of gravelly soils with fines content of less than 5% using 36 shear wave velocity test (V_s) data from the 1906 San Francisco earthquake, 1983 Borah Peak earthquake, 1989 Loma Prieta earthquake, 1993 Hokkaido-nansei earthquake, and 1995 Hyogo-ken Nanbu earthquake. Later, the simplified technique was adjusted to account for the influence of void ratio and GC on the liquefaction of gravelly soil (Chang, 2016). Following the enhancement of the data by the dynamic penetration test (DPT) and shear wave velocity test in the 2008 Wenchuan earthquake, an increasing number of models, such as the fundamental procedures (Cao and Yuan, 2010; Yuan and Cao, 2011), logistic regression (LR) models (Cao et al., 2011; Cao et al., 2013), and artificial neural network (ANN) models (Kang et al., 2014) were developed. However, the historical DPT data utilized to develop these approaches and models came from a single earthquake (the 2008 Wenchuan earthquake), therefore the models' generalization abilities need to be tested further using additional historical data. Despite their reliable and accurate results, most algorithms are difficult to apply in reality due to their extensive training and modeling procedures, as well as their "black box" aspects. Decision tree algorithms have been successfully applied to numerous geotechnical problems, such as pillar stability (Ahmad et al., 2021) and soil liquefaction potential (Ahmad et al., 2019a). When using tree algorithms to evaluate seismic gravelly soil liquefaction potential, outputs should be discrete values such as "yes" or "liquefied", "no" or "non-liquefied."

Artificial intelligence (AI) techniques have been widely used to solve real-world problems in the last 10 years, particularly in civil engineering. AI techniques have been successfully applied to a wide range of real-world scenarios, paving the way for a number of promising opportunities in civil engineering and other fields such as environmental (Froemelt et al., 2018), geotechnical and geological (Momeni et al., 2014; Armaghani et al., 2017; Mikaeil et al., 2018a; Mikaeil et al., 2018b; Ahmad et al., 2019a; Ahmad et al., 2019b; Dormishi et al., 2019; Ahmad et al., 2020a; Ahmad et al., 2020b; Ahmad et al., 2020c; Ahmad et al., 2020d; Noori et al., 2020; Ahmad et al., 2021; Ahmad et al., 2022; Amjad et al., 2022; Xie et al., 2022; Yan et al., 2022), and other sciences (Hajihassani et al., 2014; Guido et al., 2020; Morosini et al., 2021; Asteris et al., 2022) including seismic gravelly soil liquefaction potential evaluation (Kang et al., 2014). These studies introduced new ideas and methods for assessing the seismic liquefaction potential of gravelly soils. This field, however, is still being researched. The main purpose of this research is to develop new decision tree models for predicting seismic gravelly soil liquefaction potential. The decision tree algorithms have the advantage of dealing with the classification problem, making it a rational choice in classification and decision-making. The main applications of the decision tree algorithms—C4.5, random tree (RT), and logistic model tree (LMT)—have mostly been used in geotechnical engineering to predict pillar stability, slope stability, and liquefaction susceptibility (Ahmad et al., 2019a; Ahmad et al., 2021; Li et al., 2022), but critical review of existing literature suggests that despite the successful implementation of LMT, RT, reduced error pruning tree (REPT), and random forest (RF) in various domains, their implementation to predict seismic gravelly soil liquefaction potential is scarcely explored. Furthermore, one of the main significance of the developed decision tree models is that in these models, there is no need to add functional parameters (such as cyclic stress ratio in the LR model proposed by Cao et al. (2013) values affecting parameters; all parameters can be put into the model as they are, without any normalization or calibration.

2 Sample library and correlation analysis

The existing soil liquefaction case histories data are collected as supportive data for the establishment of prediction models. In this study, the dynamic penetration test (DPT) data of gravelly soil liquefaction to 234 from 17 historical earthquakes were ascertained from Hu (2021a). Most of these cases have been reviewed, screened, and corrected to result in robust database and recently referred by Hu (2021a). The seismic gravelly soil liquefaction case history data is shown in Table 1 (the complete database is available in Supplementary Appendix Table SA1), where X1 indicates moment magnitude (M_w), X2 indicates the epicentral distance (R), X3 indicates the bracketed duration (t), X4 indicates the peak ground acceleration (PGA), X5 indicates gravel content (GC), X6 indicates fines content (FC), X7 indicates average particle size (D_{50}), X8 indicates overburden stress-corrected dynamic penetration test blow count (N'_{120}), X9 indicates vertical effective overburden stress (σ'_v), X10 indicates depth to the water table (D_w), X11 indicates thickness of the impreable capping layer (H_n), and X12 indicates thickness of the unsaturated zone between groundwater table and capping layer (D_n). These twelve parameters have been widely accepted, among the researchers such as (Yuan and Cao, 2011; Hu, 2021b) as their values are relatively easy to be obtained and suitable set to evaluate seismic gravelly soil liquefaction potential. The summary of descriptive statistics of the input paramaters (i.e., M_w, R, t, PGA, GC, FC, D_{50}, N'_{120}, σ'_v, D_w, H_n, and D_n) are given in Table 2. The mean or mode imputation method is a simple and widely used method for replacing missing values (Batista and Monard, 2003). This approach is used in this

TABLE 1 Seismic gravelly soil liquefaction history data.

S. No.	X1	X2 (km)	X3 (s)	X4	X5 (%)	X6 (%)	X7 (mm)	X8	X9 (kPa)	X10 (m)	X11 (m)	X12 (m)	Liquefied?
1	7.9	96.30	40	0.21	9	9	0.5	16.01	32	1.5	1.5	0	Yes
2	7.9	94.00	40	0.24	5	53	6.15	10.79	49	0.8	0.8	0	Yes
3	7.9	95.00	40	0.24	4.9	50	5.9	20.91	46	1	1.1	0	Yes
…	…	…	…	…	…	…	…	…	…	…	…	…	
232	7.9	99.00	40	0.18	4.9	54.9	7.57	14.84	63	2	0.6	1.4	No
233	7.9	129.60	110	0.2	4.9	30	1.7	16.61	100	4.1	3.5	0.6	No
234	7.9	84.42	105	0.43	4.9	90	71.2	18.21	99	4	4	0	No

TABLE 2 Descriptive statistics of each parameter.

Parameter	X1	X2 (km)	X3 (s)	X4	X5 (%)	X6 (%)	X7 (mm)	X8	X9 (kPa)	X10 (m)	X11 (m)	X12 (m)
Max	9.2	456.38	150	0.84	45.3	90	71.2	62.12	276.8	11.7	10.65	7
Min	6.4	9.2	3.5	0.074	0	0.4	0.15	2.36	11.8	0	0	0
Mean	7.7	89.84	48.06	0.353	8.082	45.7	8.647	14.98	80.9	2.564	2.261	0.873
SD	0.713	110.372	37.136	0.156	6.130	18.397	11.191	8.862	50.778	2.026	1.483	1.217
Kurtosis	−0.334	3.677	−0.193	0.240	6.734	0.146	13.944	4.370	1.807	3.238	5.700	8.896
Skewness	0.529	2.144	0.866	0.597	2.046	−0.257	3.357	1.702	1.374	1.682	1.489	2.739

TABLE 3 Correlation coefficients between various parameters.

Parameter	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	1											
X2	0.7186	1										
X3	0.5599	0.3112	1									
X4	−0.3044	−0.3412	−0.0137	1								
X5	−0.1271	−0.1056	−0.1635	0.2267	1							
X6	−0.1485	−0.2132	0.0287	−0.0408	−0.2901	1						
X7	−0.0069	−0.1112	0.1591	−0.1057	−0.2208	0.7574	1					
X8	0.0396	−0.0636	0.2041	−0.0631	−0.1679	0.1899	0.2619	1				
X9	0.3639	0.4205	0.1964	0.0136	0.2910	−0.1501	−0.1220	−0.0683	1			
X10	0.1565	0.2438	0.2219	0.1352	0.1863	−0.0161	0.0005	−0.0667	0.6871	1		
X11	0.2049	0.2391	0.1566	−0.0480	0.0281	0.0413	0.0626	−0.0740	0.3325	0.0912	1	
X12	−0.0571	−0.0204	−0.0375	0.1335	−0.0157	0.0959	0.0017	0.0214	0.2304	0.4763	−0.3123	1

study. For each of these parameters in the considered database, the minimum (Min) and maximum (Max) limits, standard deviation (SD), kurtosis, and skewness values have been tabulated. A lower SD number indicates that the results are mainly close to the mean (*PGA*, M_w, D_n, H_n, D_w, *FC*, and N'_{120}), while a larger SD suggests a greater spread out (*R*, σ'_v, *t*, *GC*, and D_{50}) (Edjabou et al., 2017). Skewness (value might be positive, zero, negative, or undefined) assists in evaluating the extent of asymmetry of the probability distribution in the case of a real-valued arbitrary parameter from the perspective of its average value (Sharma and Ojha, 2020). Furthermore, kurtosis is typically between −10 (heavy-tailed) and +10 (light-tailed), which aids in determining the form of a probability distribution, as explained by Brown and Greene (Brown and Greene, 2006). The kurtosis values for M_w and *t* are negative and range between −0.3 and −0.1 (follow mesokurtic distribution), whereas the rest are positive values (follow leptokurtic distribution) (Benson, 1993; Lee and Ahn, 2019).

TABLE 4 Confusion matrix of binary problem.

Actual	Predicted	
	Yes	No
Yes	TP	FN
No	FP	TN

TP,true positive; TN, true negative; FP, false positive; FN, false negative.

Pearson's correlation coefficient (ρ) was used to measure the strength of a linear association between two variables. Given a pair of random variables (p, q), the following equation is used to calculate ρ:

$$\rho(p,q) = \frac{\mathrm{cov}(p,q)}{\sigma_p \sigma_q} \quad (1)$$

where cov denotes covariance, σ_p denotes the standard deviation of p, and σ_q denotes the standard deviation of q. $|\rho| > 0.8$ represents a strong correlation between m and n, values between 0.3 and 0.8 represents a moderate relationship, and $|\rho| < 0.30$ represents a weak relationship (van Vuren, 2018). Table 3 depicts the strength of the relationship between the various parameters in order of moderate to weak linear correlation. The correlation coefficient has a maximum absolute value of 0.7574, meaning there is no "strong" linear correlation. Correlation analysis is commonly used in prediction modeling to find potential predictors of an outcome variable. Correlation analysis can also be used to identify potential sources of multicollinearity in our predictor variables. When two or more predictor variables are significantly associated with each other, multicollinearity occurs. We can identify which variables contribute to multicollinearity by evaluating the correlation matrix of our predictor variables.

3 Decision tree algorithms

3.1 Random tree

Random trees comprise a forest of predictor trees. Random tree is an algorithm halfway between a simple decision tree and a random forest. The classification mechanisms include the following: The random tree classifier classifies the input vector of characteristic with each tree in the forest and then outputs the class label with the most "votes" (Witten et al., 2011).

A random tree is one that is randomly created from a set of possible trees, each of which has K random attributes at each node. In this context, "at random" indicates that any tree in the set has an equal chance of being chosen for sampling. The tree distribution is "uniform." Rapidly constructing random trees and integrating them with large sets of random trees typically yields accurate models. In recent years, there has been extensive research on random trees in the field of machine learning. This model employed the random tree approach in order to achieve the highest level of accuracy in its numerous classifier parameters, such as MinNum value–a minimum number of instances, depth–maximum depth of the tree, and seed–randomly selecting attributes, K value–number of sets utilized for randomly chosen attributes. The Decision Tree must be basic and compact for improved classification. Otherwise, the level of precision will be diminished. To obtain the highest level of precision, a random tree algorithm modifies the depth, seed, and K value. To determine the maximum parameter value, one parameter was held constant while the other was adjusted to determine the parameter with the highest accuracy.

3.2 Reduced error pruning tree

The Reduced Error Pruning Tree (REPT) is a fast decision tree learning algorithm method that combines Reduced Error Pruning (REP) and the Decision Tree (DT) (Quinlan, 1987). When the output of a decision tree is large, the DT is used to simplify the modeling process using training dataset, and the REP is used to reduce complicity of the tree structure (Mohamed et al., 2012). The pruning process in the REPT algorithm addresses the problem of backward overfitting (Quinlan, 1987). Based on the post-pruning method, the REPT algorithm attempts to find the minimal version of the most accurate sub-tree (Esposito et al., 1999; Chen et al., 2009). This model's performance is based on information gain from entropy or variance reduction and error pruning techniques (Srinivasan and Mekala, 2014). The complex decision trees can lead to overfitting and make a model less interpretable, REP helps to reduce complexity by removing the DT structure's leaves and branches (Quinlan, 1987; Galathiya et al., 2012; Mohamed et al., 2012; Pham et al., 2019).

3.3 Random forest

Breiman (2001) developed the Random Forest (RF) classifier and may be characterized as a collection of classification trees in which each tree votes on the class assigned to a given sample, with the most frequently occurring answer winning the vote (Sun and Schulz, 2015). The RF method has demonstrated its ability to handle high-dimensional data and is relatively resistant to overfitting (Breiman, 2001). This algorithm is widely used in various domains of civil engineering, including geotechnical engineering. Before model training, two parameters must be selected: the number of predictors considered at each fork of the tree and the number of random trees constructed during model construction. This machine learning has various advantages, including great performance with complicated datasets utilizing tiny calibrating and the ability to handle with high noise factors. The bagging approach is always used in a random forest to randomly select variables from the entire dataset for model calibration.

3.4 Logistic model tree

Logistic Model Tree combines the C4.5 algorithm (Quinlan, 1992) and Logistic Regression (LR) functions. The information gain ratio technique is utilized to divide the tree into nodes and leaves, and the LogitBoost algorithm (Landwehr et al., 2005) is used to fit the logistic regression functions at each node of the tree. Because it is the quickest approach for giving dependable classification accuracy, the C4.5 algorithm employs the entropy methodology for feature selection (Lim et al., 2000). The CART technique, which prunes the tree for modeling the training dataset,

TABLE 5 Definition and formulation of performance measures.

Parameter	Definition	Formulation
Acc	Rate of correctly classified instances from total instances	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$
Prec	Rate of correct predictions	$Prec = \frac{TP}{TP+FP}$ *or* $\frac{TN}{TN+FN}$
Rec	True positive rate	$Rec = \frac{TP}{TP+FN}$ *or* $\frac{TN}{TN+FP}$
F-score	Used to measure the accuracy of the experiment	$F-score = \frac{2 \times Prec \times Rec}{Prec+Rec}$
Mcc	measure the difference between the predicted classes and actual classes	$Mcc = \frac{TP \times TN - FN \times FP}{\sqrt{(TP+FP)(TN+FP)(TN+FN)(TP+FN)}}$
AUC	The capacity of a classifier to discriminate between classes and is used to summarise the *ROC* curve	-

Acc may represent how many samples are correctly identified, but it cannot demonstrate how many liquefied sites are correctly detected. Therefore, additional performance measures, such as *Prec* and *Rec*, are necessary to further evaluate the performance of a model. *Prec* and *Rec* is a pair of contradictory measures. Generally, *Prec* is large while *Rec* is not large, or *vice versa*. Therefore, a compromised index, *F-score*, is used to balance them. *Mcc* takes values in the interval [−1, 1], with "1" showing a complete agreement, "−1" a complete disagreement, and "0" presenting that the prediction was uncorrelated with the ground truth. The *Mcc* value is regarded to be the best evaluation measure for the overall performance of a classifier method (Baldi et al., 2000). *F-score* combines precision and recall values to attain a harmonic mean. *F-score* has ranged from 0 (worst value) to 1 (best value). The *AUC*, is employed to summarize the *ROC*, curve; *ROC*, curve gives five degrees of rating (Bradley, 1997): excellent (0.9–1), good (0.8–0.9), fair (0.7–0.8), poor (0.6–0.7), and not discriminating (0.5–0.6). Concisely, model having good *Acc*, larger *AUC*, high *F–score*, and high *Mcc* concurrently depicting ideal model as class imbalance are not simply eluded in soil liquefaction study.

overcomes the overfitting problem, which is a significant challenge in LMT modeling (Shahabi et al., 2013). The *IGR* can be expressed as follows:

$$\text{Gain ratio}(a) = \frac{\text{gain}(a)}{\text{split info}(a)} \tag{2}$$

where gain (a) is the information obtained after attribute an is chosen as a test for training sample classification and split info (a) is the information obtained after categorizing x training samples into n subsets (Quinlan, 1993).

The LogitBoost algorithm then conducts additive Logistic Regression with least-squares fit for each class C_i (yes or no) using the equation below (Doetsch et al., 2009):

$$L_c(x) = \sum_{i=1}^{CF} \alpha_i x_i + \alpha_0 \tag{3}$$

where $L_c(x)$ is the least-squares fit, and CF, α_i are, respectively, the number of liquefaction potential conditioning factors and the coefficient of the ith element of vector x. The posteriori probabilities in the leaves of the LMT are calculated using the linear Logistic Regression model (Landwehr et al., 2005):

$$p(c \mid x) = \frac{\exp(L_c(x))}{\sum_{c'=1}^{c} \exp(L_{c'}(x))} \tag{4}$$

where c is the number of liquefaction classes and $L_c(x)$, the least-squares fit, is transformed in such a way that $\sum_{c'=1}^{c} L_c(x) = 0$

4 Performance measures

The accuracy (*Acc*), Matthews correlation coefficient (*Mcc*), precision (*Prec*), recall (*Rec*), *F-score*, and area under the receiver operator characteristic (*ROC*) curve (*AUC*) were used to evaluate the model's performance. The performance metrics, together with their formulations and definitions, based on the confusion matrix (Table 4) described in Table 5.

5 Results and discussion

The manner in which data is divided into training and test sets has a significant impact on the results of data mining techniques (Javadi et al., 2006). The optimal parameter configuration is used to fit the prediction model to the training set, and the test set is used to evaluate model performance based on overall prediction outcomes and prediction ability for each class. Finally, the best model is chosen by comparing the overall performance of various models. If the model's prediction performance is satisfactory, it can be used for deployment. The methodological framework of tree-based classifier for predicting seismic gravelly soil liquefaction potential models is shown in Figure 1.

The entire calculation process is performed in Waikato Environment for Knowledge Analysis (WEKA) software (Version 3.9.6), a java-based and open-source application, trained the decision tree models. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. The details of the different parameters of classifiers used for the implementation of developed models in WEKA framework are summarized in Table 6. First, the search range of different parameters values is specified. In particular, for different algorithms, the search range of the same parameters is kept consistent. Further on, according to the maximum average accuracy, the optimal values for each set of parameters are obtained, which are indicated in Table 6. Based on the same dataset, these algorithms with optimal hyperparameters were then used to predict seismic gravelly soil liquefaction potential. Several performance measures based on a confusion matrix are made using training and testing datasets for gravelly soil liquefaction potential were used in order to quantify the performance measures of the proposed models. The performance results of the proposed models were obtained and compared with each other based on the same training and testing datasets. Subsequently, the confusion matrix of each model was determined, as shown in Table 7.

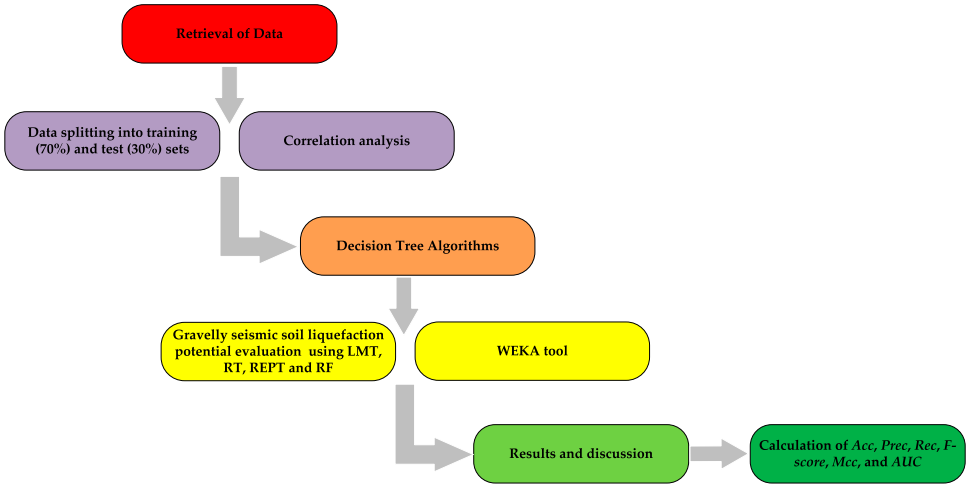


FIGURE 1
Methodological framework of Tree-based classifier for predicting seismic gravelly soil liquefaction potential.

TABLE 6 Classifiers' parameters.

LMT		RT		REPT		RF	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
fastRegression	True	KValue	0	initialCount	0	bagSizePercent	9
minNumistances	15	allowUnclassifiedInstances	False	maxDepth	−1	maxDepth	1
numBoostingIterations	−1	maxDepth	0	minNum	2	numExcutionSlots	1
splitOnResiduals	False	minNum	1	numFolds	4	numFeatures	0
useAIC	False	numFolds	2	seed	4	numIterations	1,000
weightTrimBeta	0	seed	5	spreadIntialCount	False	seed	1

TABLE 7 Confusion matrices results for training and testing datasets.

Dataset	Actual	Model							
		LMT		RT		REPT		RF	
		Predicted							
		Yes	No	Yes	No	Yes	No	Yes	No
Training	Yes	89	5	86	8	87	7	88	6
	No	12	56	14	54	12	56	24	44
Testing	Yes	39	3	38	4	34	8	39	3
	No	9	21	8	22	5	25	11	19

The values on the main diagonal indicated the number of cases correctly predicted. The *Acc*, *Prec*, *Rec*, *F-score*, *Mcc*, and *AUC* were calculated, which were listed in Table 6. To make a fair comparison, all the models are developed by applying them to the same seismic gravelly soil liquefaction case history training and testing data sets.

Figure 2, displays the bar plot of the yes and the no classes of the seismic gravelly soil liquefaction for the training and test phases. The analysis of the *Acc* together with *Rec*, *Prec*, *F-score*, *Mcc*, and *AUC* for the seismic gravelly soil liquefaction potential data set demonstrates that the LMT achieved a better prediction performance in training set succeeded

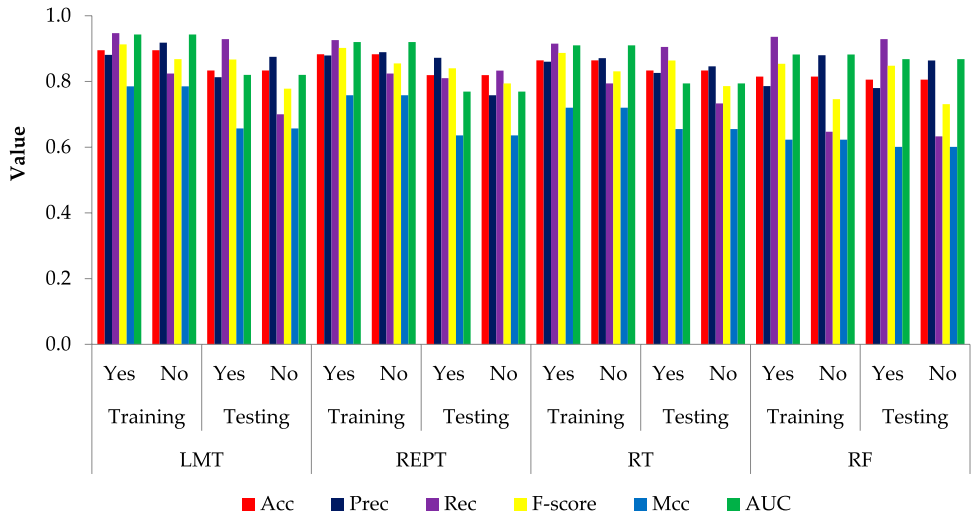


FIGURE 2
Comparison of results of developed DT models.

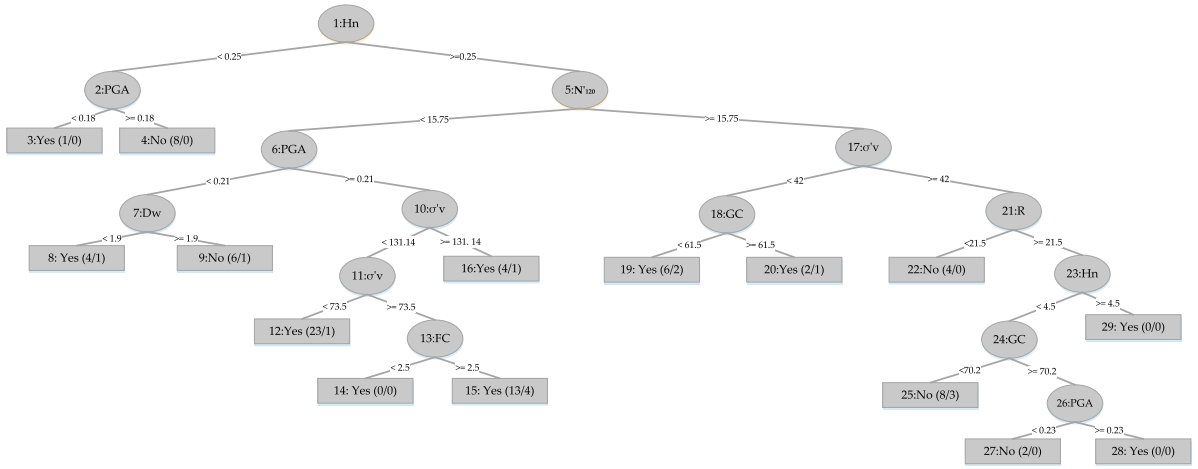


FIGURE 3
RT model.

by the REPT model, RT model and the RF model. Similarly, in test set, the LMT also achieved better prediction succeeded by the REPT model, RT model, and the RF model (see Figure 2). It can be seen that most of the cases, i.e., 145 were accurately classified using the LMT in the training dataset whereas the performance of the LMT and RT models are at par, i.e., 60 cases in the test set. Decision trees algorithms, i.e., RT and REPT are quite transparent, and are white box models that are more intuitive and interpretable than ones with other model, i.e., logistic regression and ANN models for seismic gravelly soil liquefaction potential proposed in the literature. Due to tree-like structures, the proposed models can not only obtain accurate classification results, but can also show the internal mechanism for classification results. Figure 3 shows the results of implementing RT. The size of this tree is 29 with the number of nodes being 14 and 15 leaves. The leaves of the tree represent the predictive rules of the tree. The process time of building this tree is 0.02 s. Similarly, the size of REPT is 11 with the number of nodes being 5 and 6 leaves is shown in Figure 4. The process time of building REPT is 0.01 s. The overall accuracy of LMT model based on the training and test sets were found better than the RT model. By comprehensively analyzing the *Acc* together with *Rec*, *Prec*, *F-score*, *Mcc*, and *AUC*, the rank of overall prediction performance was LMT>REPT>RT>RF. According to the *Rec*, *Prec*, *F-score* of each yes and no levels, the prediction performance for yes level was better than that for no level. The graphical output of the RT and REPT models are presented in Figures 3, 4 respectively. The numbers in parentheses at each leaf node, represent the total number of instances and the number of incorrectly classified cases. It is clear that some instances are misclassified in some leaves. The number of misclassified instances is specified after a slash. In order to create the most accurate RT and REPT

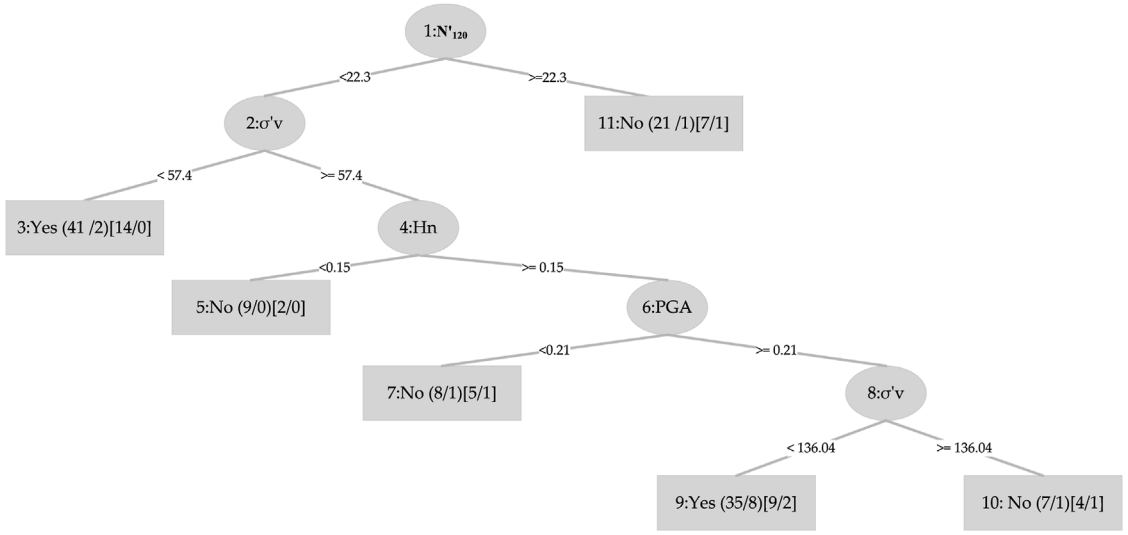


FIGURE 4
REPT model.

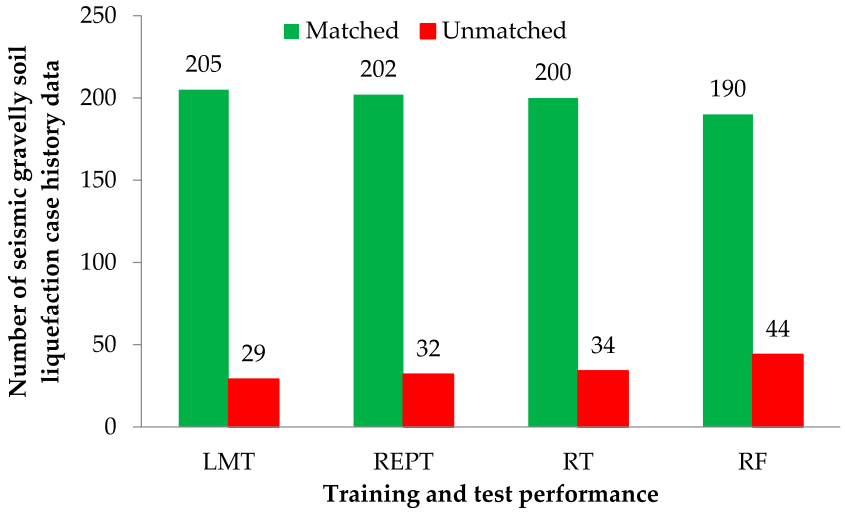


FIGURE 5
Comparison of results obtained from different models.

models, the optimal values for the minimum number of instances per leaf in WEKA were obtained through trial and error. The developed RT and REPT models, Figures 3, 4 can be used by geotechnical engineering professionals with the help of a spreadsheet to evaluate the gravelly soil liquefaction potential for a future seismic event without going into complexities of model development.

Furthermore, Figure 5 shows that the LMT was able to achieve excellent results with the lowest number of unmatched cases. For LMT, REPT, RT, and RF models, the matched and unmatched numbers were 205 and 29, 202 and 32, 200, and 34, and 190 and 44, respectively, indicating the LMT model's superiority over the other models used in this study. The error rate throughout the testing phase was low, illustrating the LMT model's high performance. It was determined that the model with the best performance for seismic gravelly soil liquefaction was the LMT, and that it could be utilized in this field for the same purpose of minimizing the associated risk.

Although the proposed models produce desired predictions, it has some limitations, such as a limited and class imbalanced dataset. In general, a limited dataset will have an impact on model generalization and reliability. While the generated models perform well with limited data sets, with *Acc* greater than 80%, prediction performance on a larger dataset should be improved. Furthermore, the dataset is class imbalanced because yes (liquefied) instances outnumber no (non-liquefied) cases. As a result, it is critical to make a larger and more balanced seismic gravelly soil liquefaction database.

TABLE 8 Rank analysis of the develop model outcomes for training dataset.

Model statistical parameter		LMT	REPT	RT	RF
Acc	Value	0.895	0.883	0.864	0.815
	Score	4	3	2	1
Mcc	Value	0.785	0.758	0.720	0.623
	Score	4	3	2	1
AUC	Value	0.945	0.920	0.910	0.882
	Score	4	3	2	1
	Total	12	9	6	3

For determining the effectiveness of developed models and comparing their robustness, rank analysis method is used. The statistical parameters are used to assign the score value in this study, with their ideal values serving as the benchmark. It depends on how many models are used. The greatest score is given to the best performing results model, and *vice versa*. The ranking ratings for two models with the same outcomes may be the same. The score attained by LMT is the highest in the training phase (12), followed by REPT (9) and RT (6) and RF (3) as presented in Table 8.

6 Conclusion

In this paper, prediction models were developed by using decision tree models such as LMT, REPT, RT, and RF for the seismic gravelly soil liquefaction potential and compared the model performances. Totally 234 case history data were used for the study with twelve different input parameters for seismic gravelly soil liquefaction potential were selected as the input variables: M_w, R, t, PGA, GC, FC, D_{50}, N'_{120}, σ'_v, D_w, H_n, and D_n. The predictive performance of the proposed models is verified and compared. In this study, the LMT model successfully achieved a high level of modeling prediction efficiency to REPT model, RT model and RF model in the training and test sets. Because all models were developed using the same methods (with the same training and test data sets), the LMT model performed the best and highest in this aspect. By comprehensively analyzing the *Acc*, *Prec*, *Rec*, *F-score*, *Mcc*, and *AUC* for yes and no classes, the LMT performed better than REPT, RT, and RF models in the training and test sets. However, in this study, the RF was deemed the lowest-performing model. For LMT, REPT, RT, and RF models, the matched and unmatched numbers were 205 and 29, 202 and 32, 200 and 34, and 190 and 44, respectively, indicating the LMT model's superiority over the other models used in this study. Furthermore, rank values of LMT, REPT, RT and RF are 12, 9, 6, and 3, respectively and the error rate throughout the testing phase was low, illustrating the LMT model's high performance. It is believed that the limitation on achieving more successful results is due to the limited number of data (234 case history data) and it is thought that the success rates of different decision tree models will increase if the data set is expanded in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

MA: Conceptualization, Methodology, Software, Writing—Original draft preparation. BTA: Methodology, Data curation, Writing—Original draft preparation. AH: Visualization, Investigation. AM: Investigation, Conceptualization. MA: Software, Validation. MMSS: Methodology, Formal analysis, Resources, Funding acquisition. RAA: Investigation, Project administration, Validation, Writing—Reviewing and Editing. MRBI: Investigation, Methodology, Writing—Reviewing and Editing.

Funding

The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program "Priority 2030" (Agreement 075-15-2021-1333 dated 30.09.2021).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1105610/full#supplementary-material

References

Ahmad, M., Al-Mansob, R. A., Kashyzadeh, K. R., Keawsawasvong, S., Sabri Sabri, M. M., Jamil, I., et al. (2022). Extreme gradient boosting algorithm for predicting shear strengths of rockfill materials. *Complexity* 2022, 1–11. doi:10.1155/2022/9415863

Ahmad, M., Al-Shayea, N. A., Tang, X.-W., Jamal, A., M Al-Ahmadi, H., and Ahmad, F. (2020d). Predicting the pillar stability of underground mines with random trees and C4. 5 decision trees. *Appl. Sci.* 10, 6486. doi:10.3390/app10186486

Ahmad, M., Kamiński, P., Olczak, P., Alam, M., Iqbal, M. J., Ahmad, F., et al. (2021). Development of prediction models for shear strength of rockfill material using machine learning techniques. *Appl. Sci.* 11, 6167. doi:10.3390/app11136167

Ahmad, M., Tang, X.-W., Qiu, J.-N., and Ahmad, F. (2019a). Evaluating seismic soil liquefaction potential using bayesian belief network and C4. 5 decision tree approaches. *Appl. Sci.* 9, 4226. doi:10.3390/app9204226

Ahmad, M., Tang, X.-W., Qiu, J.-N., Ahmad, F., and Gu, W.-J. (2020b). A step forward towards a comprehensive framework for assessing liquefaction land damage vulnerability: Exploration from historical data. *Front. Struct. Civ. Eng.* 14, 1476–1491. doi:10.1007/s11709-020-0670-z

Ahmad, M., Tang, X.-W., Qiu, J.-N., Gu, W.-J., and Ahmad, F. (2020a). A hybrid approach for evaluating CPT-based seismic soil liquefaction potential using Bayesian belief networks. *J. Central South Univ.* 27, 500–516. doi:10.1007/s11771-020-4312-3

Ahmad, M., Tang, X., and Ahmad, F. (2020c). "Evaluation of liquefaction-induced settlement using random forest and REP tree models: Taking pohang earthquake as a case of illustration," in *Natural hazards-impacts* (IntechOpen: Adjustments & Resilience).

Ahmad, M., Tang, X., Qiu, J., Ahmad, F., and Gu, W. (2019b). "LLDV-A comprehensive framework for assessing the effects of liquefaction land damage potential," in *Proceedings of 2019 IEEE 14th international conference on intelligent systems and Knowledge engineering (ISKE)* (IEEE), 527–533.

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., and Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: Modeling and performance evaluation. *Appl. Sci.* 12, 2126. doi:10.3390/app12042126

Andrus, R. D., and Stokoe, K. H. (2000). Liquefaction resistance of soils from shear-wave velocity. *J. geotechnical geoenvironmental Eng.* 126, 1015–1025. doi:10.1061/(asce)1090-0241(2000)126:11(1015)

Armaghani, D. J., Mohamad, E. T., Narayanasamy, M. S., Narita, N., and Yagiz, S. (2017). Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. *Tunn. Undergr. Space Technol.* 63, 29–43. doi:10.1016/j.tust.2016.12.009

Asteris, P. G., Lourenço, P. B., Roussis, P. C., Adami, C. E., Armaghani, D. J., Cavaleri, L., et al. (2022). Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. *Constr. Build. Mater.* 322, 126500. doi:10.1016/j.conbuildmat.2022.126500

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* 16, 412–424. doi:10.1093/bioinformatics/16.5.412

Batista, G. E., and Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Appl. Artif. Intell.* 17, 519–533. doi:10.1080/713827181

Benson, C. H. (1993). Probability distributions for hydraulic conductivity of compacted soil liners. *J. geotechnical Eng.* 119, 471–486. doi:10.1061/(asce)0733-9410(1993)119:3(471)

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30, 1145–1159. doi:10.1016/s0031-3203(96)00142-2

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Brown, S. C., and Greene, J. A. (2006). The wisdom development scale: Translating the conceptual to the concrete. *J. Coll. Student Dev.* 47, 1–19. doi:10.1353/csd.2006.0002

Cao, Z., Youd, T. L., and Yuan, X. (2013). Chinese dynamic penetration test for liquefaction evaluation in gravelly soils. *J. Geotechnical Geoenvironmental Eng.* 139, 1320–1333. doi:10.1061/(asce)gt.1943-5606.0000857

Cao, Z., Youd, T. L., and Yuan, X. (2011). Gravelly soils that liquefied during 2008 Wenchuan, China earthquake, Ms= 8.0. *Soil Dyn. Earthq. Eng.* 31, 1132–1143. doi:10.1016/j.soildyn.2011.04.001

Cao, Z., and Yuan, X. (2010). Shear wave velocity-based approach for evaluating gravel soils liquefaction. *Chin. J. Rock Mech. Eng.* 29, 943–951.

Chang, W.-J. (2016). Evaluation of liquefaction resistance for gravelly sands using gravel content–corrected shear-wave velocity. *J. Geotechnical Geoenvironmental Eng.* 142, 04016002. doi:10.1061/(asce)gt.1943-5606.0001427

Chen, J., Wang, X., and Zhai, J. (2009). "Pruning decision tree using genetic algorithms," in *Proceedings of 2009 international conference on artificial intelligence and computational intelligence*, 244–248.

Chen, L., Yuan, X., Cao, Z., Sun, R., Wang, W., and Liu, H. (2018). Characteristics and triggering conditions for naturally deposited gravelly soils that liquefied following the 2008 Wenchuan Mw 7.9 earthquake, China. *Earthq. Spectra* 34, 1091–1111. doi:10.1193/032017eqs050m

Doetsch, P., Buck, C., Golik, P., Hoppe, N., Kramp, M., Laudenberg, J., et al. (2009). "Logistic model trees with auc split criterion for the kdd cup 2009 small challenge," in *Proceedings of KDD-cup 2009 competition*, 77–88.

Dormishi, A., Ataei, M., Mikaeil, R., Khalokakaei, R., and Haghshenas, S. S. (2019). Evaluation of gang saws' performance in the carbonate rock cutting process using feasibility of intelligent approaches. *Eng. Sci. Technol. Int. J.* 22, 990–1000. doi:10.1016/j.jestch.2019.01.007

Edjabou, M. E., Martín-Fernández, J. A., Scheutz, C., and Astrup, T. F. (2017). Statistical analysis of solid waste composition data: Arithmetic mean, standard deviation and correlation coefficients. *Waste Manag.* 69, 13–23. doi:10.1016/j.wasman.2017.08.036

Esposito, F., Malerba, D., Semeraro, G., and Tamma, V. (1999). The effects of pruning methods on the predictive accuracy of induced decision trees. *Appl. Stoch. Models Bus. Industry* 15, 277–299. doi:10.1002/(sici)1526-4025(199910/12)15:4<277::aid-asmb393>3.0.co;2-b

Froemelt, A., Durrenmatt, D. J., and Hellweg, S. (2018). Using data mining to assess environmental impacts of household consumption behaviors. *Environ. Sci. Technol.* 52, 8467–8478. doi:10.1021/acs.est.8b01452

Galathiya, A., Ganatra, A., and Bhensdadia, C. (2012). Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. *Int. J. Comput. Sci. Inf. Technol.* 3, 3427–3431.

Ghani, S., Kumari, S., and Bardhan, A. (2021). A novel liquefaction study for fine-grained soil using PCA-based hybrid soft computing models. *Sādhanā* 46, 113–117. doi:10.1007/s12046-021-01640-1

Guido, G., Haghshenas, S. S., Haghshenas, S. S., Vitale, A., Gallelli, V., and Astarita, V. (2020). Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm. *Sustainability* 12, 6735. doi:10.3390/su12176735

Hajihassani, M., Armaghani, D. J., Sohaei, H., Mohamad, E. T., and Marto, A. (2014). Prediction of airblast-overpressure induced by blasting using a hybrid artificial neural network and particle swarm optimization. *Appl. Acoust.* 80, 57–67. doi:10.1016/j.apacoust.2014.01.005

Hatanaka, M., Uchida, A., and Ohara, J. (1997). Liquefaction characteristics of a gravelly fill liquefied during the 1995 Hyogo-Ken Nanbu earthquake. *Soils Found.* 37, 107–115. doi:10.3208/sandf.37.3_107

Hu, J. (2021b). A new approach for constructing two Bayesian network models for predicting the liquefaction of gravelly soil. *Comput. Geotechnics* 137, 104304. doi:10.1016/j.compgeo.2021.104304

Hu, J. (2021a). Data cleaning and feature selection for gravelly soil liquefaction. *Soil Dyn. Earthq. Eng.* 145, 106711. doi:10.1016/j.soildyn.2021.106711

Javadi, A. A., Rezania, M., and Nezhad, M. M. (2006). Evaluation of liquefaction induced lateral displacements using genetic programming. *Comput. Geotechnics* 33, 222–233. doi:10.1016/j.compgeo.2006.05.001

Kang, F., Li, J., and Zhou, H. (2014). "Artificial neural network model for evaluating gravelly soils liquefaction using shear wave velocity," in *International efforts in lifeline earthquake engineering*, 608–615.

Landwehr, N., Hall, M., and Frank, E. (2005). Logistic model trees. *Mach. Learn.* 59, 161–205. doi:10.1007/s10994-005-0466-3

Lee, J. H., and Ahn, C. K. (2019). Stochastic relaxation of nonlinear soil moisture ocean salinity (SMOS) soil moisture retrieval errors with maximal Lyapunov exponent optimization. *Nonlinear Dyn.* 95, 653–667. doi:10.1007/s11071-018-4588-0

Li, N., Zare, M., Yi, C., and Jimenez, R. (2022). Stability risk assessment of underground rock pillars using logistic model trees. *Int. J. Environ. Res. public health* 19, 2136. doi:10.3390/ijerph19042136

Lim, T.-S., Loh, W.-Y., and Shih, Y.-S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Mach. Learn.* 40, 203–228. doi:10.1023/a:1007608224229

Lin, P.-S., Chang, C.-W., and Chang, W.-J. (2004). Characterization of liquefaction resistance in gravelly soil: Large hammer penetration test and shear wave velocity approach. *Soil Dyn. Earthq. Eng.* 24, 675–687. doi:10.1016/j.soildyn.2004.06.010

Mikaeil, R., Haghshenas, S. S., and Hoseinie, S. H. (2018a). Rock penetrability classification using artificial bee colony (ABC) algorithm and self-organizing map. *Geotechnical Geol. Eng.* 36, 1309–1318. doi:10.1007/s10706-017-0394-6

Mikaeil, R., Haghshenas, S. S., Ozcelik, Y., and Gharehgheshlagh, H. H. (2018b). Performance evaluation of adaptive neuro-fuzzy inference system and group method of data handling-type neural network for estimating wear rate of diamond wire saw. *Geotechnical Geol. Eng.* 36, 3779–3791. doi:10.1007/s10706-018-0571-2

Mohamed, W. N. H. W., Salleh, M. N. M., and Omar, A. H. (2012). "A comparative study of reduced error pruning method in decision tree algorithms," in *Proceedings of 2012 IEEE International conference on control system, computing and engineering*, 392–397.

Momeni, E., Nazir, R., Armaghani, D. J., and Maizir, H. (2014). Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* 57, 122–131. doi:10.1016/j.measurement.2014.08.007

Morosini, A. F., Haghshenas, S. S., Haghshenas, S. S., Choi, D. Y., and Geem, Z. W. (2021). Sensitivity analysis for performance evaluation of a real water distribution system by a pressure driven analysis approach and artificial intelligence method. *Water* 13, 1116. doi:10.3390/w13081116

Noori, A. M., Mikaeil, R., Mokhtarian, M., Haghshenas, S. S., and Foroughi, M. (2020). Feasibility of intelligent models for prediction of utilization factor of TBM. *Geotechnical Geol. Eng.* 38, 3125–3143. doi:10.1007/s10706-020-01213-9

Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H., Bui, D. T., et al. (2019). Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* 175, 203–218. doi:10.1016/j.catena.2018.12.018

Quinlan, J. R. (1992). "Learning with continuous classes," in *Proceedings of 5th Australian joint conference on artificial intelligence*, 343–348.

Quinlan, J. R. (1987). Simplifying decision trees. *Int. J. man-machine Stud.* 27, 221–234. doi:10.1016/s0020-7373(87)80053-6

Quinlan, R. 4.5 (1993). *Programs for machine learning morgan*. San Francisco, USA: kaufmann publishers inc.

Shahabi, H., Ahmad, B., and Khezri, S. (2013). Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). *Arabian J. geosciences* 6, 3885–3907. doi:10.1007/s12517-012-0650-2

Sharma, C., and Ojha, C. (2020). "Statistical parameters of hydrometeorological variables: Standard deviation, SNR, skewness and kurtosis," in *Advances in water Resources engineering and management* (Springer), 59–70.

Sirovich, L. (1996). Repetitive liquefaction at a gravelly site and liquefaction in overconsolidated sands. *Soils Found.* 36, 23–34. doi:10.3208/sandf.36.4_23

Srinivasan, D. B., and Mekala, P. (2014). "Mining social networking data for classification using reptree," in *International journal of advance research in computer science and management studies*, 2.

Sun, L., and Schulz, K. (2015). The improvement of land cover classification by thermal remote sensing. *Remote Sens.* 7, 8368–8390. doi:10.3390/rs70708368

van Vuren, T. (2018). *Modeling of transport demand–analyzing, calculating, and forecasting transport demand: By VA profillidis and GN botzoris*. Amsterdam: Elsevier, 472. $125 (paperback and ebook), eBook ISBN: 9780128115145, Paperback ISBN: 9780128115138. Taylor & Francis: 2020.

Wang, Y., and Wang, Y.-L. (2017). Liquefaction characteristics of gravelly soil under cyclic loading with constant strain amplitude by experimental and numerical investigations. *Soil Dyn. Earthq. Eng.* 92, 388–396. doi:10.1016/j.soildyn.2016.10.029

Witten, I., Frank, E., and Hall, M. (2011). *Data mining: Practical machine learning tools and techniques 3 edition*. San Francisco: Morgan Kaufmann.

Xie, C., Nguyen, H., Choi, Y., and Armaghani, D. J. (2022). Optimized functional linked neural network for predicting diaphragm wall deflection induced by braced excavations in clays. *Geosci. Front.* 13, 101313. doi:10.1016/j.gsf.2021.101313

Yan, T., Shen, S.-L., and Zhou, A. (2022). Identification of geological characteristics from construction parameters during shield tunnelling. *Acta Geotech.* 18, 535–551. doi:10.1007/s11440-022-01590-w

Yegian, M., Ghahraman, V., and Harutiunyan, R. (1994). Liquefaction and embankment failure case histories, 1988 Armenia earthquake. *J. geotechnical Eng.* 120, 581–596. doi:10.1061/(asce)0733-9410(1994)120:3(581)

Youd, T., Harp, E., Keefer, D., and Wilson, R. (1985). The Borah peak, Idaho earthquake of october 28, 1983—liquefaction. *Earthq. spectra* 2, 71–89. doi:10.1193/1.1585303

Yuan, X., and Cao, Z. (2011). A fundamental procedure and calculation formula for evaluating gravel liquefaction. *Earthq. Eng. Eng. Vib.* 10, 339–347. doi:10.1007/s11803-011-0070-4

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 27 July 2023
DOI 10.3389/feart.2023.1132890

OPEN ACCESS

EDITED BY
Zefa Yang,
Central South University, China

REVIEWED BY
Honglei Yang,
China University of Geosciences, China
Xuemin Xing,
Changsha University of Science and Technology, China

*CORRESPONDENCE
Peixian Li,
lipx@cumtb.edu.cn

SPECIALTY SECTION
This article was submitted to Geohazards and Georisks, a section of the journal Frontiers in Earth Science

RECEIVED 28 December 2022
ACCEPTED 30 March 2023
PUBLISHED 27 July 2023

CITATION
Tang W, Wang M, Li P, Wang G, Yan Y and Yan W (2023), Ground subsidence associated with mining activity in the Ningdong coal base area, northwestern China revealed by InSAR time series analysis.
Front. Earth Sci. 11:1132890.
doi: 10.3389/feart.2023.1132890

Ground subsidence associated with mining activity in the Ningdong coal base area, northwestern China revealed by InSAR time series analysis

Wei Tang[1,2,3,4], Mingliang Wang[5], Peixian Li[5]*, Guorui Wang[6], Yueguan Yan[5] and Weitao Yan[7]

[1]State Key Laboratory of Water Resource Protection and Utilization in Coal Mining, CHN Energy Shendong Coal Group Co., Ltd., Beijing, China, [2]National Institute of Clean-and-Low-Carbon Energy, Beijing, China, [3]Key Laboratory of Natural Resources Monitoring and Supervision in Southern Hilly Region, Ministry of Natural Resources, Changsha, China, [4]State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology, Beijing, China, [5]College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing, China, [6]Institute of Land and Resources Investigation and Monitoring, Yinchuan, China, [7]State Collaborative Innovation Center of Coal Work Safety and Clean-Efficiency Utilization, Henan Polytechnic University, Jiaozuo, China

Ningdong coal base area located in northwestern China is one of the largest coal-producing bases in China. The aim of this work is to investigate a regional-scale mining subsidence over the Ningdong coal base area, by using both conventional and advanced Differential Synthetic Aperture Radar Interferometry (DInSAR) methods. Fifteen L-band SAR images from ALOS-2 satellite and 102 C-band images from Sentinel-1A satellite spanning between November 2014 and July 2019 were used for the analysis. To increase the spatial extent of the displacement signal because of decorrelated effects, we modified the traditional Small Baseline Subset (SBAS) method to incorporate the coherence into the inverse problem, hereafter we call it coherence-based SBAS method. Instead of excluding decorrelated pixels present in the interferograms, we keep all the pixels in the time series analysis and down-weighted the decorrelated pixels with coherence. We performed the coherence-based SBAS method to both the two SAR datasets to obtain the subsidence rate maps and displacement time-series over the mining areas, and compared the results with that from the traditional stacking InSAR method. We evaluated the effectiveness of L-band and C-band DInSAR for monitoring mining subsidence by comparing differential interferograms and displacements derived from SBAS method between ALOS-2 and Sentinel-1A data. Compared to C-band, L-band SAR are less affected by phase aliasing due to large displacement gradients. The most significant subsidence was found at Maliantai mine with −264 mm/year detected by SBAS method from Sentinel-1 data. We validated the InSAR displacement accuracy by comparing both ALOS-2 and Sentinel-1 results with 18 GPS stations above five active mining regions. The average RMSE between InSAR and GPS measurements is 28.4 mm for Sentinel-1 data and 21 mm for ALOS-2 data. Our results demonstrate that the combined exploitation of L-band and C-band SAR data through both conventional and advanced DInSAR methods could be crucial to monitor ground subsidence in mining areas, which provides insights into subsidence dynamics and determine the characteristic surface response to longwall advance.

KEYWORDS
mining subsidence, Ningdong coal base, DInSAR, stacking, SBAS

1 Introduction

Land subsidence is one of the most important hazards associated with underground coal mining activities. Surface deformation due to excavation or destruction of mining operations may cause serious damages to urban structures and infrastructures within the mining area. The impacts of surface displacement depend on several factors including the type and size of the mine, the type and size of the structure, and the magnitude of displacement itself. Surface displacement is an indicator of cave development and stress redistribution, and short-period, high-resolution monitoring has the potential to gain insight into displacement dynamics and determine the characteristic surface response to underground mining activities (Wempen, 2020). The northwestern China is rich in coal resources and has good mining conditions, coal production has been transferred to the western region of China (Cai et al., 2022). Longwall mining is a particularly important method of coal production in the western China because coal reserves are deep. When ground subsidence is planned as part of a longwall mining development, a subsidence control plan is required. As a caving method, longwall mining can be used in deep mines with high recovery efficiency. Monitoring the spatial distribution and temporal evolution of land subsidence is critical for mining design, therefore mining activity is implemented to ensure subsidence occurs under predictable and control.

Mining subsidence was traditionally monitored using geodetic methods, including leveling and Global Positioning System (GPS), and occasionally differential analysis using Digital Elevation Model (DEM) derived from Light Detection and Ranging (LiDAR) (Zhu et al., 2020; Liu et al., 2022). Space-borne Differential Synthetic Aperture Radar Interferometry (DInSAR) has been effective to monitor ground surface displacement with the advantages of high accuracy, wide coverage, and independence of weather. Over the mining areas, this technique has proven to be particularly useful to monitor mining subsidence (Wegmuller et al., 2000; Przylucka et al., 2015; Ilieva et al., 2019). However, the traditional DInSAR technique based on a single image pair is vulnerable to decorrelated effects and atmospheric delays. InSAR time series analysis methods have been proposed to exploit the redundancy in InSAR multitemporal stacks to mitigate noise and nuisance signals such as atmospheric effects, residual topography, and decorrelation. While many InSAR time series analysis methods have been developed in the last 2 decades, most of them share similar characteristics which can be categorized into two types of technique (Ferretti et al., 2001; Berardino et al., 2002; Hooper et al., 2007): Persistent Scatterer Interferometry (PSI) and Small Baseline Subset (SBAS).

Subsidence observation to comply with mining regulations is one of the reasons for monitoring subsidence. However, the presence of fast subsidence prevents the InSAR from a robust unwrapping of the interferogram due to phase aliasing and temporal decorrelated effects. A combination of conventional and advanced DInSAR has been proposed to monitor fast evolving mining subsidence over Bytom City, Poland (Przylucka et al., 2015). By comparison of L-band and X-band interferograms over mining areas in central Utah, it is found that L-band data are less affected by phase aliasing effects due to large displacement gradients and by temporal decorrelation due to changes in the surface conditions over time (Wempen and McCarter, 2017). With a short revisit time of 6 days from Sentinel-1 constellation and a prediction model based on the Knothe-Budryk theory, it is possible to track the mining deformation life cycle (Ilieva et al., 2019). A model-assisted phase unwrapping approach was proposed to tackle the phase aliasing problem for the cases when the subsidence gradient exceeds the maximum measurable gradient of DInSAR technique (Dai et al., 2021). Recently, a deep convolutional neural network has been proposed to detect and map localized, rapid mining subsidence from wrapped interferograms, and a phase unwrapping network is designed to unwrap the cropped interferogram patches centered on the detected subsiding locations (Wu et al., 2022).

In this paper, we analyzed 15 L-band SAR images from ALOS-2 satellite and 102 C-band images from Sentinel-1A satellite spanning between November 2014 and July 2019 to investigate land subsidence over a large coal mining area in Ningdong coal base, northwestern China. In order to suppress noise from decorrelated pixels, we modified the traditional Small Baseline Subset (SBAS) method to incorporate the coherence into the inverse problem (we call it coherence-based SBAS method). The signals in the decorrelated pixels over the mining region are down-weighted in the SBAS analysis by using the coherence as weighting matrix. The stacking and coherence-based SBAS InSAR time series analysis methods were applied to analyze the two SAR datasets to obtain the subsidence rate maps and displacement time-series over the mining areas. We evaluated the capability of L-band and C-band SAR data in extracting rapid mining subsidence by comparing the interferograms and displacement time series. We validated the InSAR displacement time-series by comparing both ALOS-2 and Sentinel-1 results with 18 continuous GPS stations above five active mining regions in this coal base. Our results provide insights into subsidence dynamics and determine the characteristic surface response to longwall advance and support the hazard mitigation in the coal base.

2 Methodology

2.1 Conventional DInSAR processing

InSAR exploits the phase difference between two SAR images taken from slightly different sensor positions and extracts information about the earth's surface. The computed

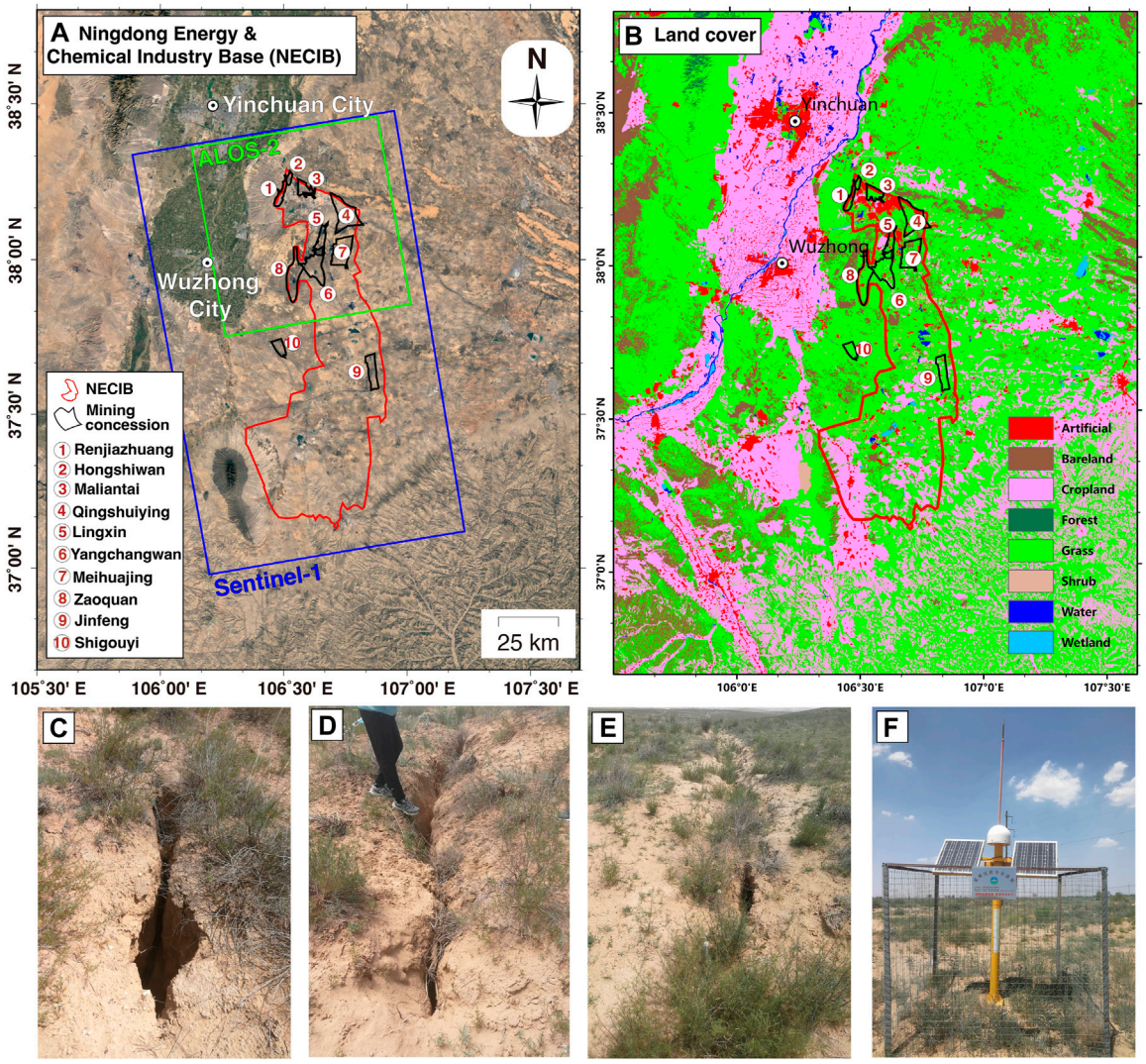


FIGURE 1
(A) Location of the Ningdong Energy and Chemical Industry Base (NECIB), as outlined by the red polygon. The ten mining concessions located in the region are outlined by black polygons and numbered from 1 to 10 in the figure. The names of the corresponding mining concession are marked in the legend. The green rectangle indicates the coverage of the ALOS-2 image and the blue rectangle indicates the processing region of the Sentinel-1 image. **(B)** The 30-m resolution land cover in the study area, which is obtained from GlobeLand30 (Chen et al., 2014). **(C–E)** The mining induced ground fissures in the NECIB **(F)** One of the continuous GPS stations in the NECIB.

interferogram contains phase variation $\Delta\varnothing_{int}$ from several contributing factors (Hanssen, 2001), most importantly the flat-earth phase $\Delta\varnothing_{flat}$, the topographic phase $\Delta\varnothing_{DEM}$, atmospheric conditions $\Delta\varnothing_{atm}$, and other noise $\Delta\varnothing_{n}$, and lastly the eventual surface displacement $\Delta\varnothing_{disp}$ which occurred between the two image acquisition dates. Thus, the phase components of an interferogram can be written as:

$$\Delta\varnothing_{int} = \Delta\varnothing_{DEM} + \Delta\varnothing_{flat} + \Delta\varnothing_{disp} + \Delta\varnothing_{atm} + \Delta\varnothing_{n} \quad (1)$$

DInSAR attempts to remove the topographic phase $\Delta\varnothing_{DEM}$ from the measured interferogram $\Delta\varnothing_{int}$ with the goal of isolating $\Delta\varnothing_{disp}$ and facilitating precise surface displacement mapping. Removing the topographic phase contribution can be achieved by using an external Digital Elevation Model (DEM). The surface displacement ΔR_{disp} in radar line-of-sight (LOS) direction can be obtained after removing the phase components of atmospheric effects $\Delta\varnothing_{atm}$ and phase noise $\Delta\varnothing_{n}$.

$$\Delta R_{disp} = -\frac{\lambda}{4\pi}\Delta\varnothing_{disp} \quad (2)$$

Longer wavelengths tend to have some penetration of vegetation, dry soils, and ice; phase measurements from longer wavelengths tend to be less sensitive to small changes in the surface conditions over time. Additionally, the maximum deformation gradients measurable by DInSAR depend significantly on the radar band and on the ground resolution (pixel size) of the image (Wempen and McCarter, 2017). Longer wavelengths are less sensitive to deformation per pixel than shorter

TABLE 1 Production capacity, mining method, and technique of the main coal mines in NECIB [adapted from (Cai et al., 2022)].

No.	Coal mines	Production capacity (10^4 t a^{-1})	Brief introduction of coal mining method and technology
1	Renjiazhuang	360	Adopting the strike longwall comprehensive mechanized fully mechanized mining method and fully-caving management of the roof
2	Hongshiwan	60	Adopting the strike longwall comprehensive mechanized mining method, fully mechanized full seam one passing mining technology, and fully-caving management of the roof
3	Maliantai	360	Adopting the strike longwall retreating mining method, fully mechanized full seam one passing mining technology, and fully-caving management of the roof
4	Qingshuiying	500	Adopting the strike longwall comprehensive mechanized fully mechanized mining method and fully-caving management of the roof
5	Lingxin	320	Adopting the strike longwall retreating mining method, comprehensive mechanized coal mining technology, and fully-caving management of the roof
6	Yangchangwan	1,200	Adopting the strike longwall comprehensive mechanized mining method, fully mechanized full seam one passing mining technology, and fully-caving management of the roof
7	Meihuajing	1,200	Adopting the strike longwall comprehensive mechanized mining method, fully mechanized full seam one passing mining technology, and fully-caving management of the roof
8	Zaoquan	800	Adopting the strike longwall comprehensive mechanized fully mechanized mining/caving mining method and fully-caving management of the roof
9	Jinfeng	400	Adopting the strike longwall comprehensive mechanized mining method, fully mechanized full seam one passing mining technology, and fully-caving management of the roof.
10	Shigouyi	90	-

TABLE 2 Parameters of SAR data used in this study.

Satellite	ALOS-2	Sentinel-1
Wavelength	L band (24.245 cm)	C band (5.546 cm)
Resolution (range × azimuth)	9.1 m × 5.3 m	2.7–3.5 m × 22 m
Swath width	70 km	250 km
Flight direction	Ascending	Ascending
Time span	29/11/2014–6/7/2019	03/12/2014–10/7/2019
Number of scenes	15	102
Small baseline thresholds (spatial × temporal)	200 m × 300 days	150 m × 50 days
Number of formed interferograms	25	211

wavelengths, and larger deformation gradients are measurable in higher resolution data.

Besides the interferometric phase, the coherence between the reference and the repeat image is estimated as an indicator for the quality of the phase information. Basically, it shows if the images have strong similarities and are therefore usable for interferometric processing. Coherence can be computed as (Hanssen, 2001):

$$\gamma = \frac{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{M} \cdot \boldsymbol{S}^{\star}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{M} \cdot \boldsymbol{M}^{\star} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{S} \cdot \boldsymbol{S}^{\star}}} \tag{3}$$

where M, S are the complex value of a pixel in the reference and repeat images and N stands for the size of a window used for coherence estimation. Loss of coherence can produce poor interferometric results and is caused by temporal (over vegetation and water bodies), geometric (errors or inaccuracies in the orbit metadata) and volumetric decorrelation (potential scattering mechanisms of voluminous structures, such as complex vegetation or dry surfaces). The accuracy of DInSAR can be limited by phase decorrelation. Low coherence indicates an increase in the noise level of InSAR phase measurements, and InSAR analysis methods typically use a coherence threshold to mask out noisy pixels. In addition, large displacement gradients can cause phase aliasing. When the phases are saturated, the absolute deformation cannot be directly measured using conventional DInSAR technique. To limit the effect of phase decorrelation and phase aliasing effects, data were paired over the shortest possible

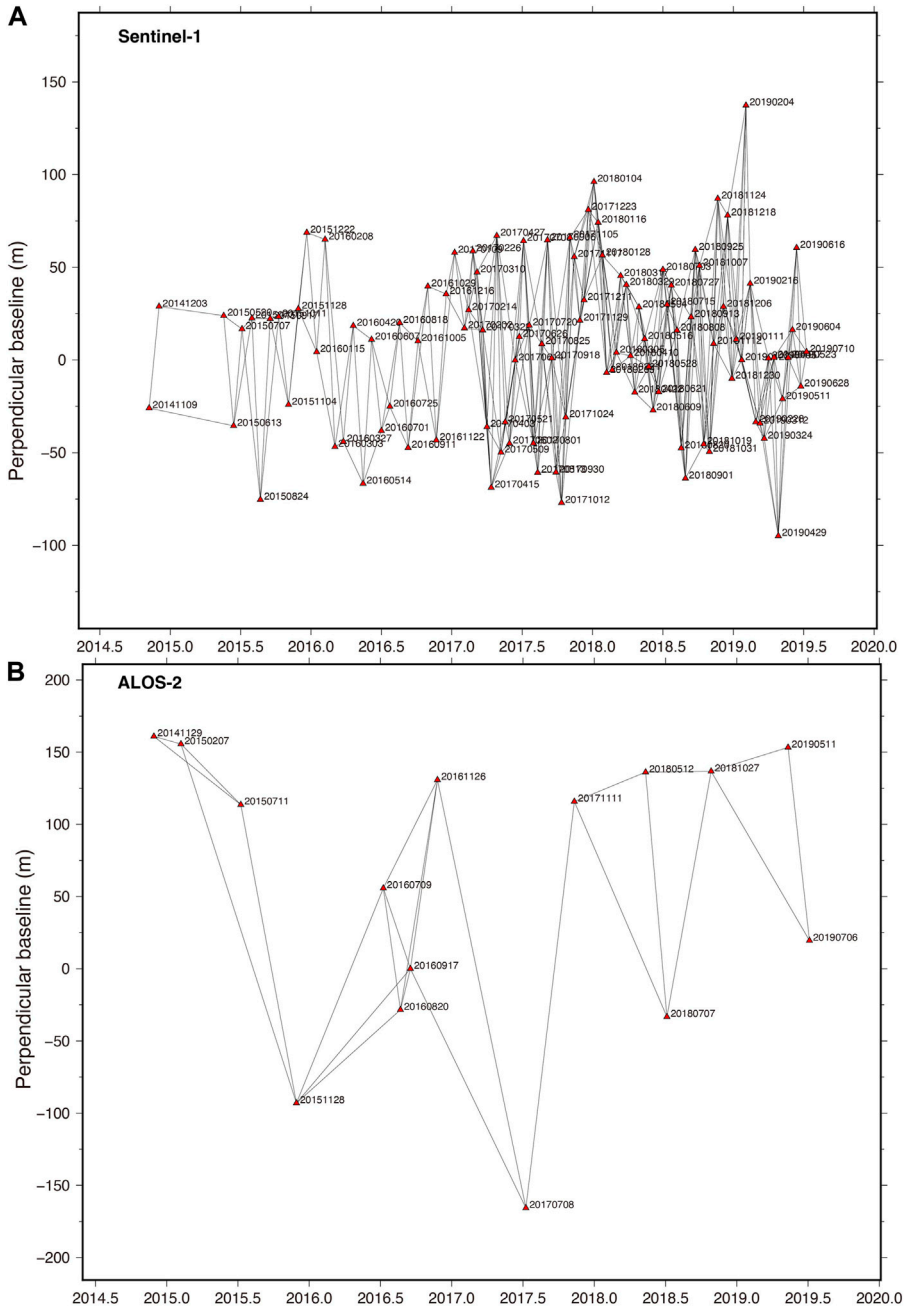


FIGURE 2
Spatial and temporal baselines for Sentinel-1 **(A)** and ALOS-2 data **(B)**. The red triangles indicate acquisition dates, and the lines represent interferometric pairs. A total number of 211 interferograms were generated from Sentinel-1 data and 25 interferograms were formed from ALOS-2 data.

periods, and the data were accumulated, additively, to estimate the subsidence over longer periods.

2.2 InSAR stacking method

To minimize the impact of atmospheric phase disturbances and unwrapping errors, a series of images can be analyzed to get a more reliable result. This is the stacking method which was originally proposed by Sandwell and Price (1998). By averaging displacements of several dates, the impact of atmospheric disturbances and unwrapping errors is reduced, and the actual displacement patterns become more evident. This assumes that deformation trend is linear between the first and the last image pair, and the phase noises are random in time. However, the assumption of linear deformation behavior might not be true in the mining region,

TABLE 3 Available continuous GPS stations in the study.

Mines	Number of monitoring stations	Site name	Longitude	Latitude	Time span
Renjiazhuang	4	RJZ-1	106.470	38.185	7/2/2015–27/5/2020
		RJZ-2	106.476	38.180	7/2/2015–27/5/2020
		RJZ-3	106.475	38.188	7/2/2015–27/5/2020
		RJZ-4	106.479	38.185	7/2/2015–27/5/2020
Maliantai	3	MLT-1	106.565	38.230	25/5/2017–27/5/2020
		MLT-2	106.563	38.257	8/3/2018–28/5/2020
		MLT-3	106.562	38.253	8/9/2017–28/5/2020
Lingxin	3	LX-1	106.639	38.050	2/6/2015–27/5/2020
		LX-2	106.637	38.047	24/5/2015–27/5/2020
		LX-3	106.636	38.044	14/6/2015–21/2/2020
Yangchangwan	4	YCW-1	106.643	37.979	12/8/2015–27/5/2020
		YCW-2	106.644	37.977	11/10/2015–27/5/2020
		YCW-3	106.645	37.975	13/11/2015–27/5/2020
		YCW-4	106.641	37.973	29/4/2015–5/1/2020
Meihuajing	4	MHJ-1	106.695	38.017	13/6/2015–27/5/2020
		MHJ-2	106.692	38.007	27/2/2016–27/5/2020
		MHJ-3	106.688	37.999	11/9/2015–6/1/2019
		MHJ-4	106.685	37.994	3/8/2015–27/5/2020

especially over the mines under active excavation. The LOS displacement rate V_{disp} can be averaged by a stack of interferograms as,

$$\boldsymbol{V_{disp}} = \frac{\boldsymbol{\lambda}}{\boldsymbol{4\pi}} \cdot \frac{\sum_{i=1}^{N} \Delta\varnothing_i \cdot \Delta \boldsymbol{T_i}}{\sum_{i=1}^{N} \Delta \boldsymbol{T_i}^2} \tag{4}$$

where $\Delta\varnothing_i$ is the differential phase of the ith interferogram. ΔT_i is the corresponding time span of the interferometric pair, which is used for the weight computation during the averaging. The standard deviation of the LOS displacement rate can be computed as (Zhang et al., 2021),

$$\boldsymbol{std_V} = \sqrt{\frac{\boldsymbol{1}}{\boldsymbol{N}} \cdot \frac{\sum_{i=1}^{N} \left(\Delta \boldsymbol{T_i}^2 \cdot \left(\frac{\Delta\varnothing_i}{\Delta T_i} - \boldsymbol{V_{disp}} \right)^2 \right)}{\sum_{i=1}^{N} \Delta \boldsymbol{T_i}^2}} \tag{5}$$

2.3 Coherence-based SBAS method

The SBAS algorithm is originally proposed by Berardino et al. (2002), which is based on an appropriate combination of interferometric pairs characterized by small spatial and temporal baselines in order to limit the decorrelation effect. The traditional SBAS method is still less effective over vegetated surfaces due to the loss of coherence; pixels with low coherence are typically excluded from the analysis, leading to the reduced spatial coverage of the final InSAR displacement map. In addition, in areas with significant gradient deformation out of the detection threshold of InSAR technique, such as over the active mining areas, the maximum mining subsidence can be up to several meters; the traditional InSAR time series methods commonly underestimate the magnitude of the large displacement field. Several variants have been developed to improve the performance of SBAS method in extracting surface displacements. One of these variants is using the metrics of phase quality (phase coherence) as a weight of the phase observations in the SBAS analysis (Schmidt and Bürgmann, 2003; Tong and Schmidt, 2016), which we refer to coherence-based SBAS method in our study.

Instead of omitting partly decorrelated pixels in the interferogram, all pixels are kept in the SBAS inverse problem and the phase observations were weighted based on the coherence for each pixel in each differential interferogram. The decorrelated pixels will be down-weighted in the time-series analysis due to their low coherence information. This coherence-based SBAS method is less sensitive to the noise in the input phase data, and it produces a spatially coherent signals with dense coverage. The weighted least squares inversion for each InSAR pixel is (Tong and Schmidt, 2016):

$$\boldsymbol{W}\begin{bmatrix} \boldsymbol{1} & \boldsymbol{1} & \boldsymbol{0} & \dots & \boldsymbol{\beta B_1} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{1} & \dots & \boldsymbol{\beta B_2} \\ \dots & \dots & \dots & \dots & \dots \\ \boldsymbol{\rho/\Delta t_1} & \boldsymbol{-\rho/\Delta t_1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\rho/\Delta t_2} & \boldsymbol{-\rho/\Delta t_2} & \dots & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{m_1} \\ \boldsymbol{m_2} \\ \dots \\ \boldsymbol{m_s} \\ \boldsymbol{\Delta h} \end{bmatrix} = \boldsymbol{W} \begin{bmatrix} \boldsymbol{d_1} \\ \boldsymbol{d_2} \\ \dots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} \tag{6}$$

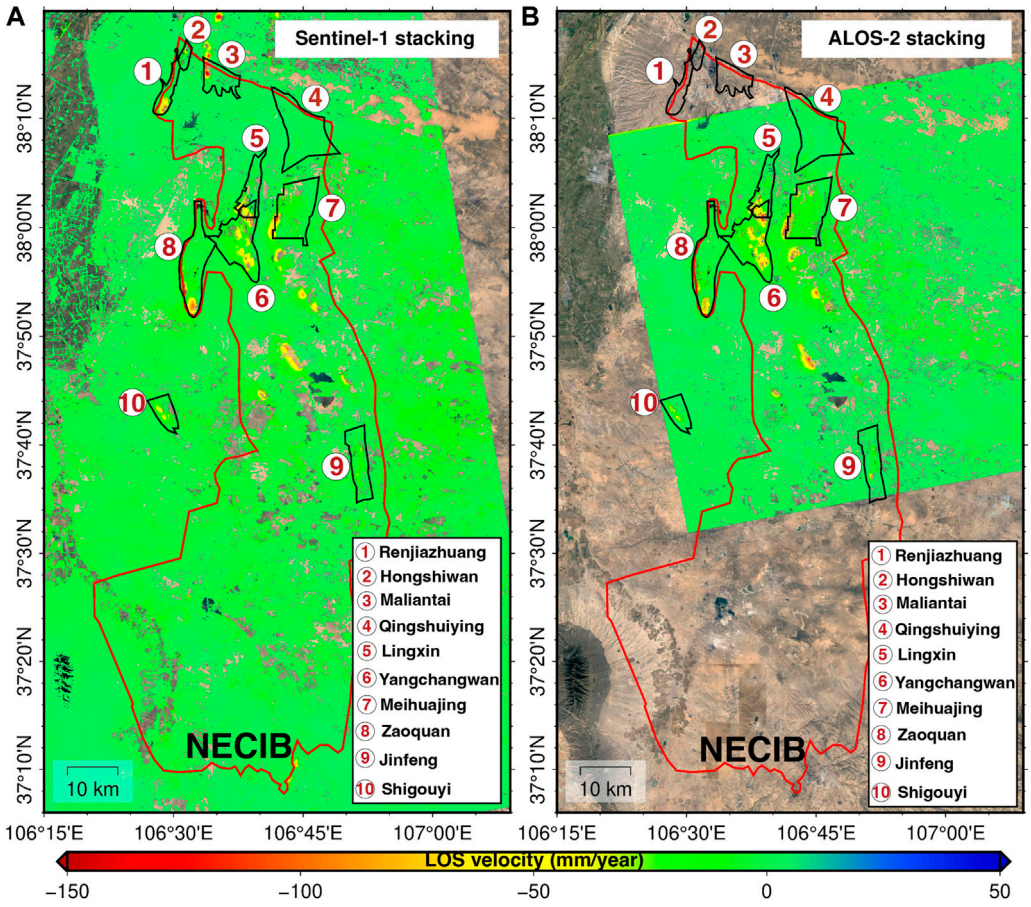


FIGURE 3
Average LOS subsidence rates derived by InSAR stacking method from Sentinel-1 **(A)** and ALOS-2 **(B)** images. The ten mining concessions in the NECIB region are outlined by black polygons and numbered from 1 to 10 in the figure.

where W is the weighting matrix and $W = diag\{\gamma_1, \gamma_2, \gamma_3, \cdots, \gamma_n\}$ with γ_i the coherence of the ith interferogram (see Eq. 3). m_j is the incremental displacement for each SAR epoch, d_i is the LOS phase of the ith interferogram, B_i is the perpendicular baseline of the ith interferogram, Δh is the DEM error, ρ is the smoothing factor to penalize the unrealistically outliers in the data. β is the scale factor determined by radar wavelength and the incidence angle of the radar signal and the distance from the radar sensor to the ground. The kernel matrix has a size of $[n \times s + 1]$, where n is the number of differential interferograms and s is the number of temporal incremental. The atmospheric correction is done in combination with the SBAS time series estimation using an the common-point-stacking method as described in (Tymofyeyeva and Fialko, 2015; Tang et al., 2022). After that, this method solves the inverse problem in Eq. 6 on a pixel-by-pixel basis to retrieve the time series of cumulative displacement as:

$$\boldsymbol{M}(\boldsymbol{t}) = \sum_{i=1}^{s} \boldsymbol{m_j} \tag{7}$$

The average displacement rates are finally calculated by a linear regression to the InSAR-retrieved displacement time series.

3 Study area and data

3.1 Description of the study area

Ningdong Energy and Chemical Industry Base (NECIB) is one of the largest coal mining areas in China. It is located in Ningxia Hui Autonomous Region, northwestern China, which covers an area of 3,484 km^2 (Figure 1A). The Ningdong coal base is located in the NECIB, which is one of the largest coal-producing bases in China. This coal base has a coal reserve of 38.6 billion tons, accounting for 84% of the total amount across Ningxia province. With the exploitation of coal resources, the surface deformation due to underground coal extraction has caused damages to the mine facilities and surrounding infrastructures and buildings (Wang et al., 2018; Wang et al., 2021). By 2019, it has been documented that the affected area of mining subsidence in NECIB was up to 22.102 km^2 and 191 ground fissures were formed (see some fissures in Figure 1) (Wang et al., 2018). At the same time, mining subsidence has destroyed the natural state of the land, caused water and soil loss, land desertification, and threatened the entire ecological environment in this region. The land cover over this

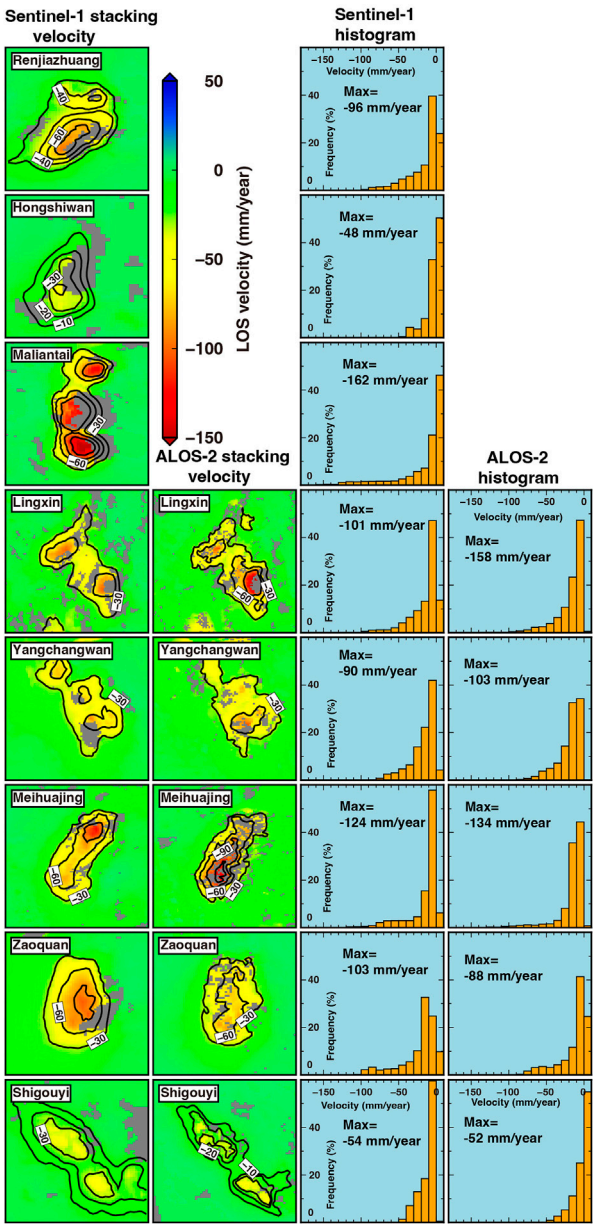


FIGURE 4
Zoomed-in view of the average LOS subsidence rates derived by the stacking InSAR method over the active mining areas. The contours of subsidence rate are superimposed on the map. Note that the contour interval varies for each velocity map. Qingshuiying and Jinfeng mines are not mapped in this figure because no significant displacement patterns were found in these two mines. The first column is velocity derived from Sentinel-1, the second column is velocity from ALOS-2, the third column is the histogram of the Sentinel-1 velocity, and the fourth column is the histogram of the ALOS-2 velocity. The ALOS-2 images do not cover Renjiazhuang, Hongshiwan, and Maliantai. The maximum subsidence rates are indicated in the third and fourth column.

region is shown in Figure 1B. The primary land use is covered by grass, which accounts for 62.3% of the territory inside the NECIB, and the second is cropland, accounting for 20.3%.

The coal mining activities in this region mainly includes 10 coalfields: Renjiazhuang, Hongshiwan, Maliantai, Qingshuiying, Lingxin, Yangchangwan, Meihuajing, Zaoquan, Jinfeng and Shigouyi, as shown in Figure 1. Mining activities started in 2003 and operated till now. Most of the coal mines have adopted the comprehensive mechanized coal mining method with the longwall mining technique, and the caving method is used to manage the roof. A few coal mines adopt the room and pillar mining method. However, it should be noted that some coal mines adopt two or more coal mining methods simultaneously. The production capacity, mining method, and technique of main coal mines in the NECIB are listed in Table 1.

Over longer periods (years), the characteristics of subsidence including the angle of draw, displacement, maximum subsidence relative to seam height, depend on both the mining and geological conditions. The mining method, the width of the mined area, the reserve recovery, the rate of mine development, the topography, and the strata thickness, depth, and dip are significant parameters that affect the magnitude, extent, and duration of subsidence. Regional hydrology, geologic structures, and the geotechnical properties of the strata also affect subsidence.

3.2 SAR data and interferograms generation

We selected SAR images from ALOS-2 and Sentinel-1 satellites for mining subsidence monitoring. Parameters of SAR data used in this study are listed in Table 2. The coverage of the SAR images can be seen in Figure 1. ALOS-2 satellite works in L band with a wavelength of 24.2 cm, which is conducive to maintaining the interferometric coherence and extracting the large surface deformation in the mining area. We obtained 15 archived acquisitions from ALOS-2 satellite, spanning the period from 29 November 2014 to 6 July 2019. The maximum theoretical detectable displacement between two fringes is 12.1 cm every 10 m (i.e., pixel resolution) for ALOS-2 image. SAR images from Sentinel-1 mission were acquired in C band with a wavelength of 5.6 cm, which is more vulnerable to the decorrelation than L-band data and is more likely to cause phase aliasing due to the rapid surface deformation. The maximum theoretical detectable displacement between two fringes is only 2.8 cm every 30 m for the Sentinel-1 image. However, Sentinel-1 data has a short sampling interval of only 12 days, and data acquisitions are regularly sampled and uniformly distributed in time, which improves the performance of the retrieval of fast deformation over mining areas. A total number of 102 images from Sentinel-1 was used in this study spanning the period from 03 December 2014 to 10 July 2019, which is many more images than that of ALOS-2. These two independent SAR datasets were both acquired in ascending orbit and in similar time spanning which makes it possible to directly compare results between the two SAR datasets.

For the coherence-based SBAS processing, we formed 211 interferograms from Sentinel-1 scenes, with a 50-day temporal baseline threshold and a 150-m perpendicular baseline threshold (Figure 2A). The pairs 20141109–20150603 and 20141203–20150520 were manually connected to avoid an isolated cluster in the interferogram combination due to the large time gap between the first two SAR acquisitions (20141109 and 20141203) and the subsequent acquisitions.

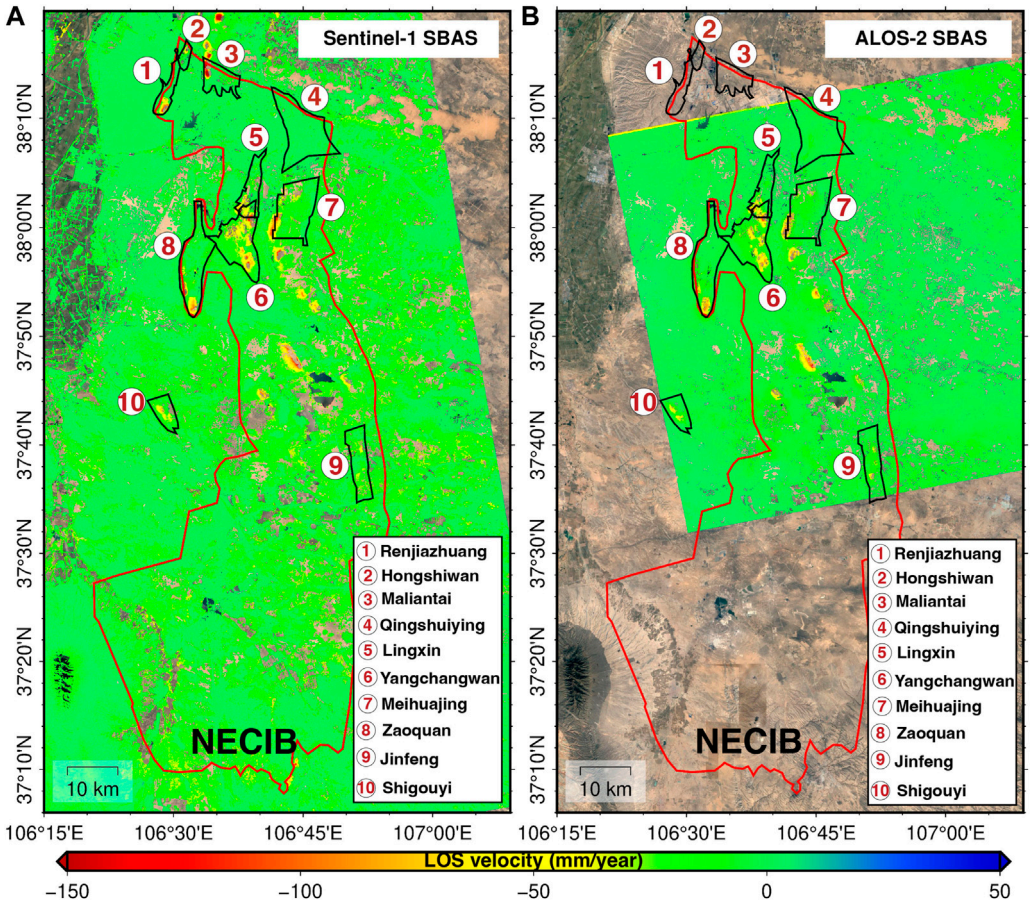


FIGURE 5
Averaged LOS subsidence rates derived by the coherence-based SBAS method from Sentinel-1 **(A)** and ALOS-2 **(B)** images.

TABLE 4 Maximum subsidence rates and maximum cumulated subsidence from ALOS-2 and Sentinel-1 data over the ten mines in NECIB from stacking and SBAS methods.

No.	Mine	Max. Subsidence rate (mm/year)				Max. Cumulated subsidence (mm)			
		Sentinel-1		ALOS-2		Sentinel-1		ALOS-2	
		Stacking	SBAS	Stacking	SBAS	Stacking	SBAS	Stacking	SBAS
1	Renjiazhuang	−96	−138	-	-	−448	−498	-	-
2	Hongshiwan	−48	−80	-	-	−223	−310	-	-
3	Maliantai	−162	−264	-	-	−755	−960	-	-
4	Qingshuiying	−61	−78	−25	−46	−284	−313	−115	−216
5	Lingxin	−101	−161	−158	−136	−471	−553	−726	−598
6	Yangchangwan	−90	−131	−103	−98	−419	−477	−473	−413
7	Meihuajing	−124	−131	−134	−116	−578	−484	−616	−506
8	Zaoquan	−103	−120	−88	−105	−480	−448	−404	−521
9	Jinfeng	−33	−53	−54	−55	−153	−205	−248	−250
10	Shigouyi	−54	−90	−52	−79	−251	−260	−239	−349

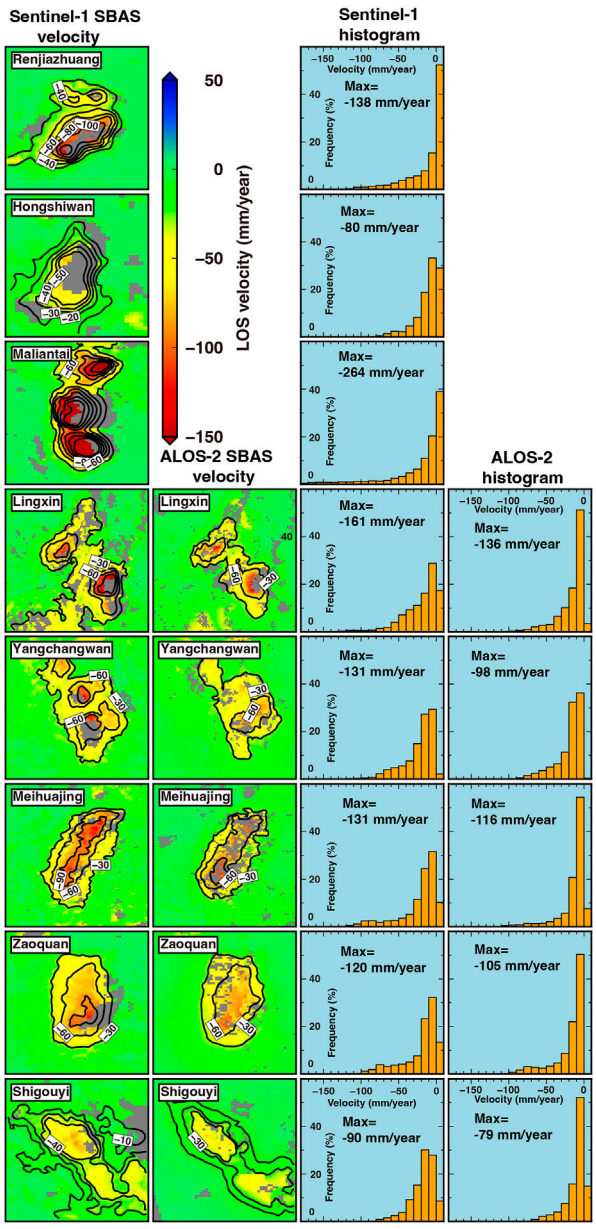


FIGURE 6
Zoomed-in view of the average LOS subsidence rates derived by the coherence-based SBAS method over the active mining areas.

C-band data from Sentinel-1 satellite are strongly decorrelated by vegetation or the fast-mining subsidence over this region. The short 50-day temporal baseline threshold was set to mitigate this effect. For the ALOS-2 data with a longer radar wavelength, we set larger thresholds for the spatial (200 m) and temporal (300 days) baselines, and 25 interferograms were constructed (Figure 2B). We also discarded interferograms with strong decorrelation by visual inspection.

All the interferograms were produced using the open source InSAR software GMTSAR (Xu et al., 2017). The software is built on a Generic Mapping Tools (GMT) foundation (Wessel et al., 2019). The ALOS-2 data are acquired in the traditional strip-map mode and the image coregistration was implemented based on the standard image cross-correlation method, which achieves sufficient accuracy for the interferometric processing (~1/10 pixel). While for the Sentinel-1 data, which are the Interferometric Wide Swath (~250 km) product acquired using the Terrain Observation with Progressive Scans SAR (TOPSAR) mode, the image alignment was implemented with the enhanced spectral diversity (ESD) to achieve the acceptable azimuth registration accuracy (better than 1/200 pixel) (Xu et al., 2017). The topographic phase components were removed using an existing DEM with 30 m spatial resolution from SRTM (Shuttle Radar Topography Mission) (Farr et al., 2007). For the spatial filter, we used a Gaussian filter with a 0.5 gain at a wavelength of 200 m in azimuth and ground range and the decimation factors were set to 8 in range and 2 in azimuth direction. This produces displacement results in geographic coordinate with a grid spacing of ~50 m. Phase unwrapping is performed by the software SNAPHU (Statistical-Cost, Network-Flow Algorithm for Phase Unwrapping) software (Chen and Zebker, 2001).

3.3 Continuous GPS data

In order to monitor the surface movement caused by coal mining, a local surveying institute has deployed high-precision real-time continuous GPS stations in some coal mines in the NECIB (Wang et al., 2018; Wang et al., 2021). At each mine, an independent GPS network was established with a reference station installing on stable ground away from the active mining surface and several monitoring stations installing above the surface of the working panel. In this study, we have access to a total of 18 GPS monitoring stations located in five mining areas, including Renjiazhuang, Maliantai, Lingxin, Yangchangwan and Meihuajing mines (Table 3). Four GPS monitoring stations are available for each mine in Renjiazhuang, Yangchangwan and Meihuajing and three stations are available for each mine in Maliantai and Lingxin. A GPS site photo can be found in Figure 1F. GPS measurements are in three dimensions: East-West (EW), North-South (NS), and Up-Down (UD), often called ENU displacements with respect to a specific reference frame or a few reference points. The GPS antenna's sampling frequency in time is 1 h. The hourly displacement time series were resampled to a daily sampling interval and were smoothed with a 7-day moving box. The GPS three-dimensional NEU displacement were projected into InSAR LOS direction to validate the accuracy of InSAR-derived displacements in Section 5.3.

4 Results and analysis

4.1 Stacking displacement velocity maps

We applied the stacking InSAR method to the unwrapped differential interferograms formed in Section 3.2 (Figure 2) and the averaged LOS displacement rates is obtained using Eq. 4, as

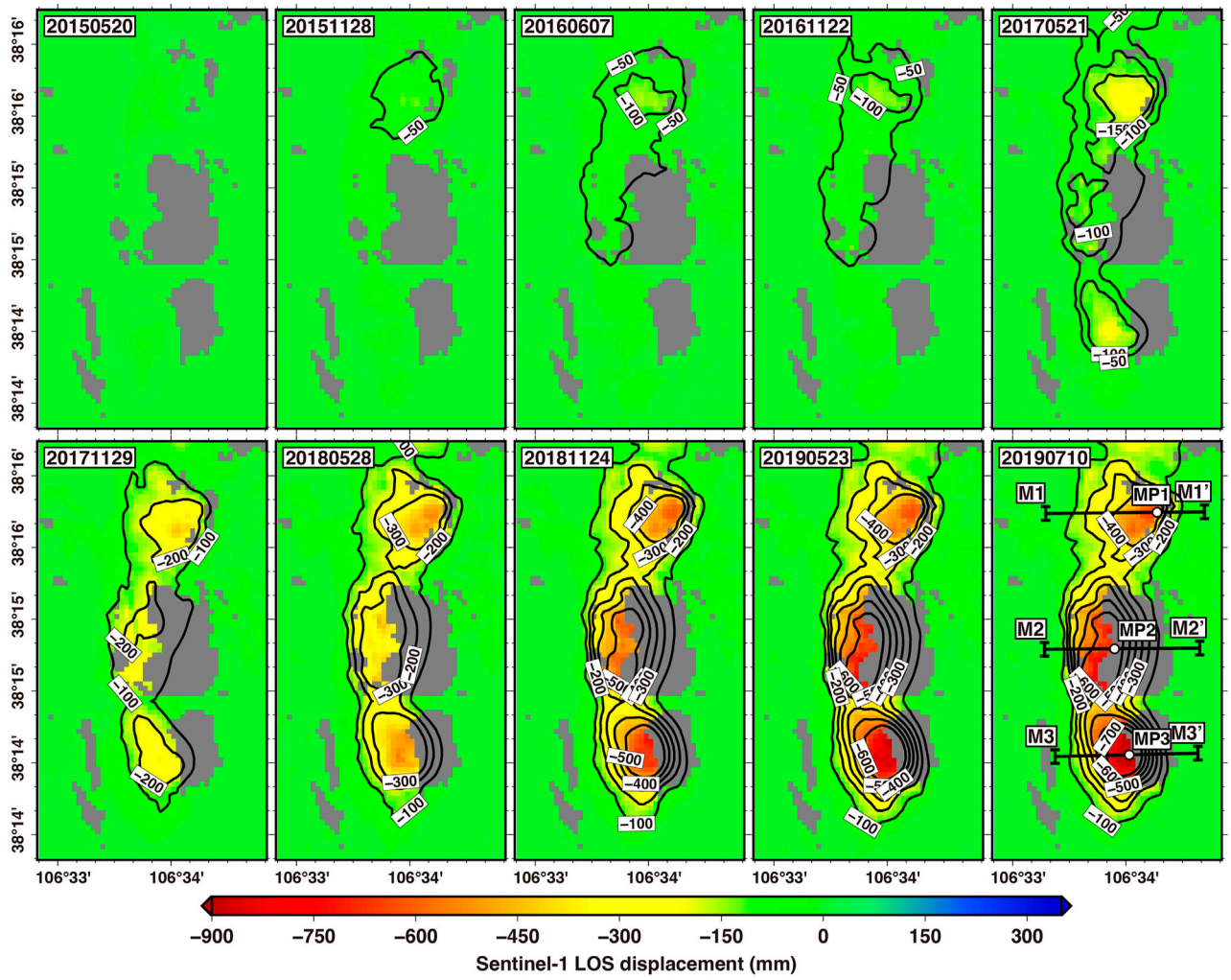


FIGURE 7
The spatial and temporal pattern of ground subsidence over Maliantai mine in NECIB, derived by the coherence-SBAS method from Sentinel-1 data. The subsidence maps are shown at intervals of about 6 months. Three subsidence bowls were formed in this mining concession region. The horizontal black lines on 20190710 (M-M1′, M2-M2′ and M3-M3′) represent the locations of the profiles across the three subsidence bowls, which are shown in Figure 8. The white points (MP1-MP3) on 20190710 marks the location of the displacement time-series shown in Figure 8. These points are located on each profile with the maximum cumulative subsidence. The subsidence contours are superimposed on each map. All the displacements are referenced to the first acquisition date (20141203) of Sentinel-1 data.

shown in Figure 3. In the results, pixels with temporally averaged coherence values smaller than 0.25 were masked out. The observation of the retrieved results for the two SAR datasets (Figures 3A, B) reveals almost the same spatial distribution and magnitude of the estimated subsidence. Except the Qingshuiying and Jinfeng mines, significant subsidence over these mining areas can be clearly identified in the velocity maps, suggesting a clear relationship between the mining activities and the measured deformation. The zoomed-in view of the averaged LOS subsidence rates from both Sentinel-1 and ALOS-2 datasets and velocity distribution histograms in these active mining areas are shown in Figure 4. The most significant subsidence was identified in Maliantai mine, with a maximum subsidence rate up to 162 mm/year in the LOS direction, which is related to the ongoing mining activities. The least significant subsidence occurred in the mines of Hongshiwan and Shigouyi, with a maximum rate of only ~50 mm/year, which was probably associated with the residual mining subsidence after mine closure. However, we also found several subsidence features outside of the boundary of the ten mining concessions, such as in the south of Yangchangwan and Meihuajing mines. This might be due to mining activities belonging to another mining concessions which are not included in our study. From Figure 4, we can see that the spatial pattern of the mining subsidence and the distribution of subsidence rates (see histograms in Figure 4) are similar for the results from Sentinel-1 and ALOS-2. Theoretically, the ALOS-2 data with a longer radar wavelength in L band can detect a larger mining subsidence rate than that of Sentinel-1 data, which can be manifested in the observations in mines of Lingxin, Yangchangwan and Meihuajing (see Figure 4).

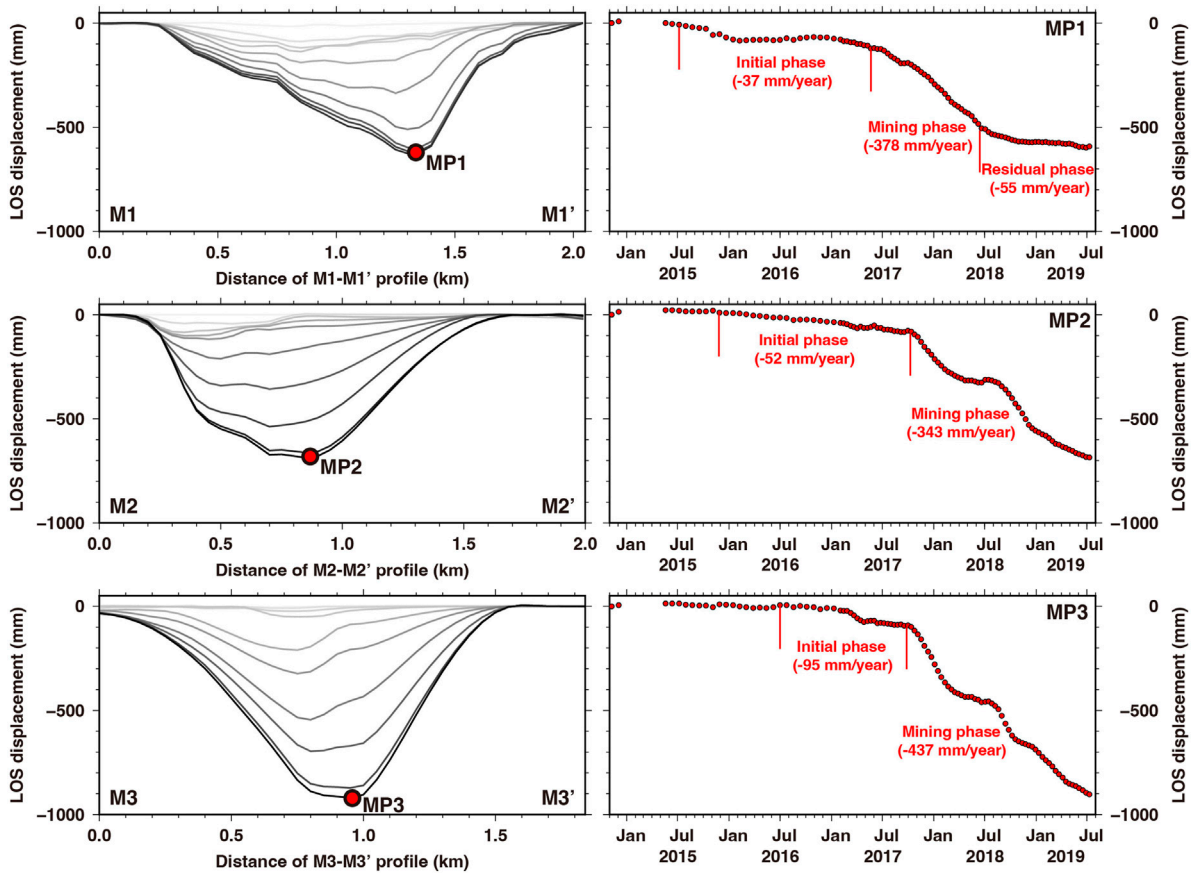


FIGURE 8
Progressively cumulated subsidence along profiles M1-M1′, M2-M2′ and M3-M3′ across the three subsidence bowls over Maliantai mine (left panel). The locations of the three profiles can be found in Figure 7. The red points (MP1-MP3) on each map represent the locations where the maximum cumulated subsidence resides along each profile (see the point locations in Figure 7), and the displacement time-series of each point is shown in the right panel.

4.2 SBAS displacement velocity maps

The unwrapped differential interferograms formed in Section 3.2 were also analyzed with the coherence-based SBAS method and the resulting LOS subsidence rate maps are shown in Figure 5. The coherence map of each interferogram were used as the weighting matrix in the computation (see Eq. 6). Generally, the SBAS results in Figure 5 are very similar to the average LOS subsidence rates obtained by the stacking method in Figure 3. We also plotted the zoomed-in view of the average LOS subsidence rates derived from SBAS method and their velocity distribution histograms in the active mining areas, as shown in Figure 6. The maximum subsidence rates and cumulated subsidence over all the ten mines in the NECIB were listed in Table 4. Comparing the results of Sentinel-1 and ALOS-2 data in Figure 6, we found that the Sentinel-1 data detected higher subsidence rates than the ALOS-2 data. We believed that the maximum detectable displacement can be improved by the short revisiting time (12 days) and regularly temporal sampling acquisitions of the Sentinel-1 data. In contrast to Sentinel-1 acquisitions, ALOS-2 images were sparsely and irregularly acquired in time, leading to more severe phase aliasing and decorrelation effects in the interferometric pairs with longer temporal baselines. When comparing the displacement rates between stacking and SBAS methods from the same SAR sensor (Figures 4, 6), we observed that the rates derived from SBAS method are generally larger than that derived from stacking method, especially for the Sentinel-1 results. The largest discrepancy was found at the active mining area of Maliantai that the maximum subsidence rate derived by SBAS method is 264 mm/year from Sentinel-1 data (Figure 6) while the stacking method largely underestimated the subsidence with the maximum rate of only 162 mm/year (Figure 4). This is because the stacking method is assuming a linear behavior of the temporal evolution of surface subsidence and simply averaging the cumulative displacement to obtain the subsidence rates, which does not hold for the mining-induced subsidence, especially during the active mining period. For the coherence-based SBAS method, the estimates of surface motion do not assume a linear model in the analysis. However, we found that over all the ten mines, the SBAS-

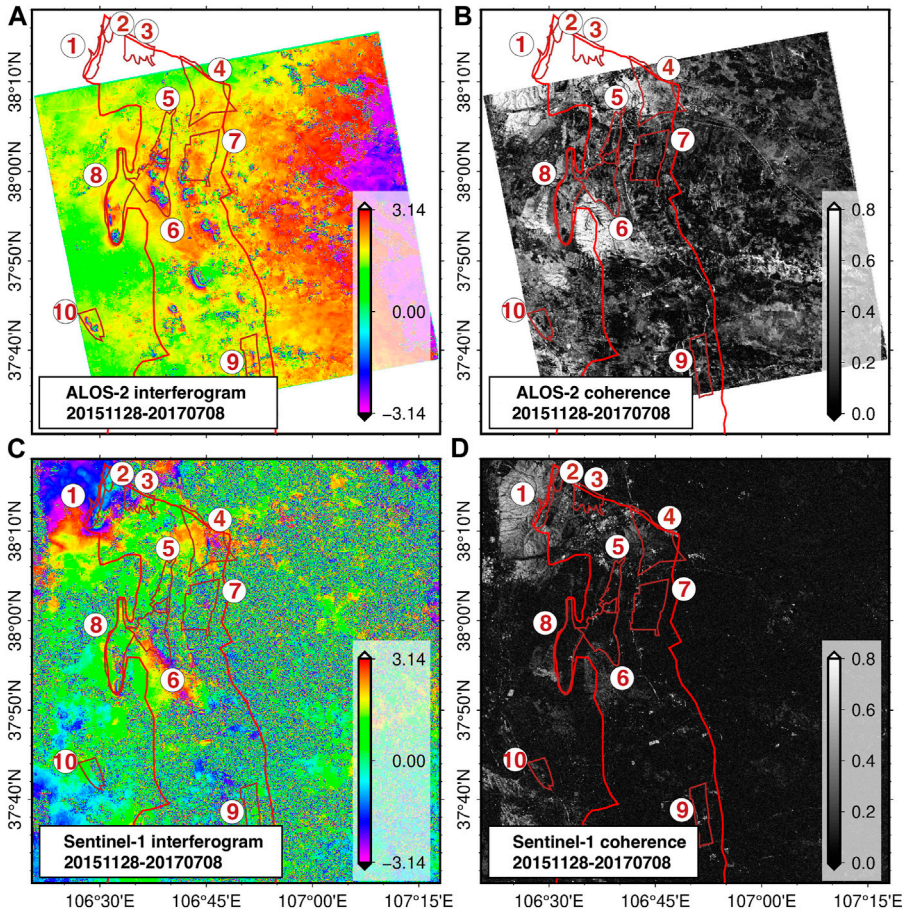


FIGURE 9
Examples of differential interferograms and the corresponding coherence maps from ALOS-2 and Sentinel-1, processed with GMTSAR. The time interval (588 days) of the reference and repeated scene of the interferometric pair is the same for the two interferograms. The coherence map is computed using Eq. 3.

derived rates from ALOS-2 are not always larger than the stacking-derived rates. This is again most likely due to the sparse sampling and irregular distribution in time domain of the ALOS-2 acquisition.

4.3 Spatial pattern and temporal evolution of active mining subsidence

In contrast to the stacking InSAR method, which can only retrieve the averaged displacement rate, the SBAS method can obtain displacement time series at each SAR acquisition date. In this section, we show the spatial distribution and temporal evolution of the ground displacements over the active mine of Maliantai in NECIB (Figure 7). Subsidence over the mine of Maliantai shows the largest rate reaching 264 mm/year in the LOS direction from Sentinel-1 data during the time period from December 2014 to July 2019 (Figure 6), indicating an ongoing active mining activities in this mine. The results of the SBAS method help to better understand the kinematics of subsidence over the mine (Figure 7). The results show that three separate subsidence bowls were formed in this mining concession region, of which the southern subsidence bowl has the largest cumulated subsidence up to 960 mm. The middle subsidence bowl also shows a significant subsidence with a maximum cumulated subsidence up to 785 mm. Due to the fast subsidence rates with a large spatial gradient, significant loss of coherence can be observed over these two mines, especially in the east side of the subsidence center (missing coherent pixels). The subsidence bowl in the north shows the mildest subsidence and the maximum cumulated subsidence is 650 mm. More coherent pixels present in this mining area because a better coherence remained.

Figure 8 provides the progression of subsidence over time along profiles M1-M1′, M2-M2′ and M3-M3′ across the three subsidence bowls, from the 10 Sentinel-1 acquisition dates in Figure 7. For a complete view of the subsidence bowl along these profiles, the decorrelated areas where InSAR pixels are absent were interpolated using an adjustable tension continuous curvature spline (Smith and Wessel, 2012). The maximum measured subsidence during the study period is −600 mm along M-M1′, −687 mm along M-M2′ and 900 mm along M-M3′. Together with the displacement time series at points MP1, MP2 and MP3 (the largest cumulated subsidence point on each profile) in Figure 8, we can track the temporal evolution of

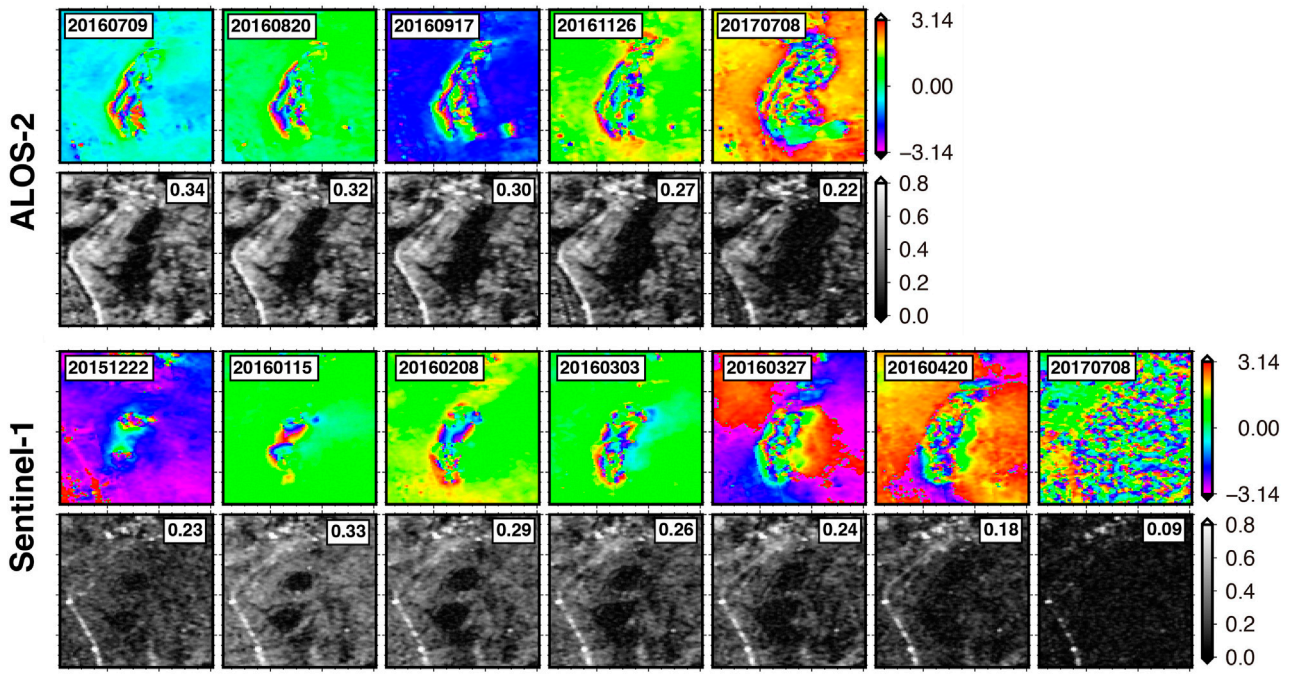


FIGURE 10
Zoomed-in view of differential interferograms and coherence maps of ALOS-2 and Sentinel-1 data over Meihuajing mine in the NECIB. A common reference image for these interferograms is set to 20151128 for both ALOS-2 and Sentinel-1 data. The repeated acquisition dates for each interferogram are marked on each map. The average coherence for each interferogram is indicated in the figure.

the mining subsidence at each mine. Three distinct phases of the surface subsidence induced by mining extraction can be clearly identified on the displacement time series at the location of MP1. The subsidence is slow during the initial phase of mining exploitation from July 2015 to July 2017, with a subsidence rate of −37 mm/year. The subsidence becomes linear over the mining phase during August 2017-July 2018 and the subsidence rate increases to −378 mm/year. Lastly, the subsidence rate decreases through the residual phase with −55 mm/year from August 2018 to the end of our study period (10 July 2019). The residual subsidence can last for decades after the end of the underneath exploitation, which is connected to the visco-plastic responses of the overburden and the compaction of caved zones. The amplitude of the late subsidence is related to the mined thickness (Modeste et al., 2021). Over the middle subsidence bowls, the subsidence rate at the locations of MP2 during the initial phase between January 2016 and October 2017 is −52 mm/year. During the mining phase between November 2017 to the end of our study period, the subsidence rate increases to −343 mm/year at point MP2. The initial phase of the mining subsidence in the south is between July 2016 and October 2017 and the subsidence rate is −95 mm/year at point MP3 during this period. The rate increases to −437 mm/year during the mining phase from November 2017 to the end of our study period. We did not see a decreasing subsidence rate at the two locations MP2 and MP3 during our study period, indicating that underground mining activities were ongoing in the mines. However, we found a pattern of relative stable ground between April 2018 and July 2018 for both MP2 and MP3, which might be due to a stoppage stage of coal mining in this 4-month period.

5 Discussion

5.1 Comparison of differential interferograms between Sentinel-1 and ALOS-2

To compare subsidence obtained from Sentinle-1 and ALOS-2 dataset, we first show examples of differential interferograms from the two SAR sensors in Figure 9. From this comparison we also illustrate the effects of phase aliasing due to fast subsidence and coherence loss over the active mining areas, which is one of the limitations of the conventional DInSAR technique. The reference and repeated images of the ALOS-2 and Sentinel-1 interferometric pairs have the same acquisition dates (reference acquisition is 20151128 and repeated acquisition is 20170708), so the phase in the two interferograms contains the same magnitude of displacement. Very local circular shaped fringes induced by mining subsidence were clearly identified in the ALOS-2 interferogram (Figure 9A). The locations of these fringes correspond with the locations of coal mining areas, indicating a clear relationship between the mining activities and measured displacement. Most of mining-induced fringes cannot be identified from the Sentinel-1

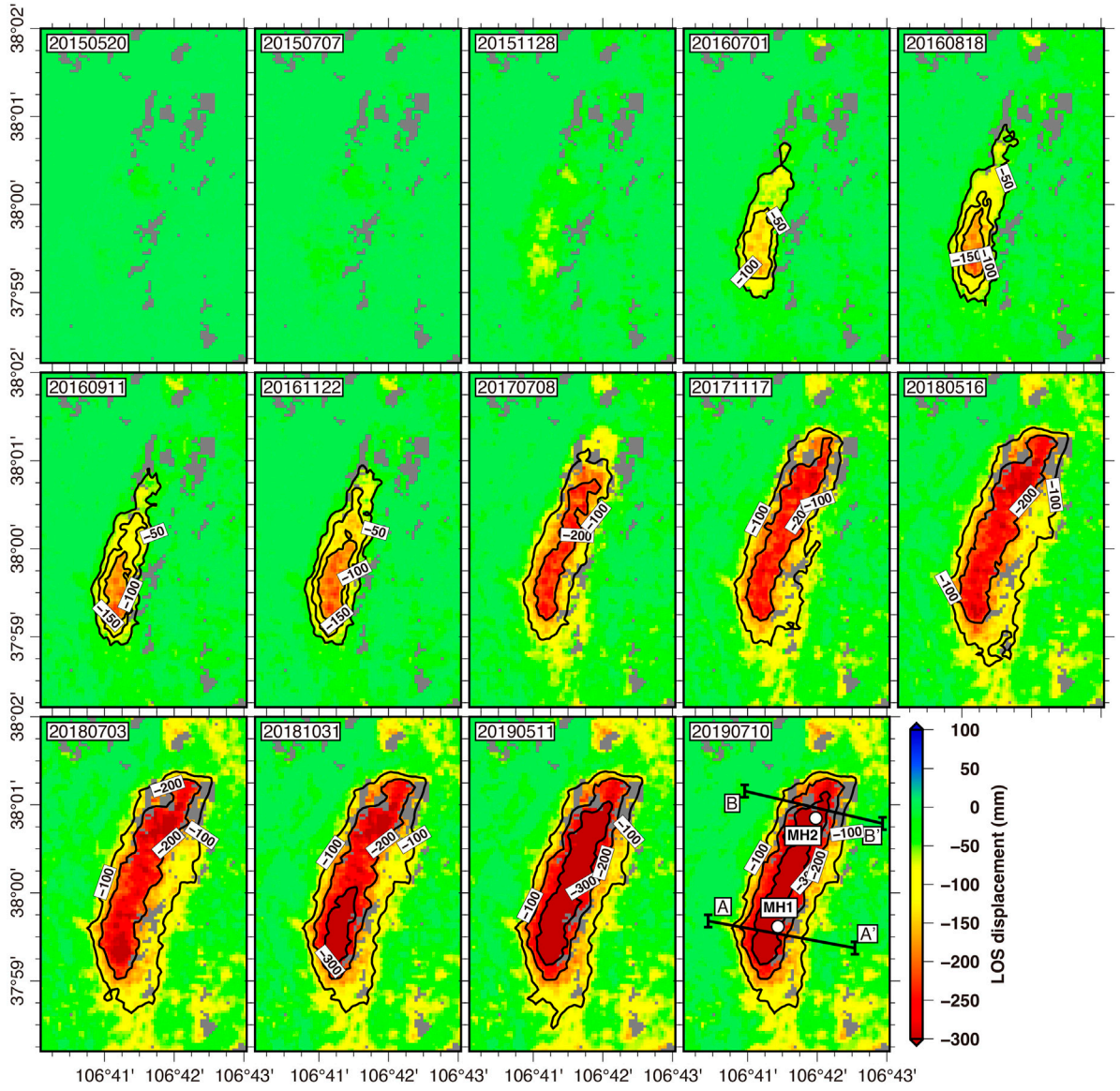


FIGURE 11
The progressive accumulation LOS subsidence over Meihuajing mining area acquired from Sentinel-1 data. For comparison, we only showed the 14 Sentinel-1 acquisition dates closest to ALOS-2 acquisition dates. The black lines on 20190710 (A-A′ and B-B′) are the profiles across the mining subsidence basin, which are shown in Figure 13. The white points (MH1 and MH2) on 20190710 marks the location of the displacement time-series shown in Figure 13.

interferogram (Figure 9C) because the C-band data is more vulnerable to large gradient deformation. Coherence loss (Figures 9B–D) can be observed for both the two interferograms due to the long-time interval (588 days) between the reference and the repeated acquisition. The coherence loss increased over the active mining areas due to the large gradient deformation, causing the surface scattering characteristics change significantly between the two acquisition dates. Obviously, the L-band ALOS-2 interferogram generally remains better coherence than the C-band Sentinel-1 interferogram. Significant atmospheric artifacts were also found in the top left corner of Figure 9C which even obfuscates the deformation fringes in the Sentinel-1 interferogram.

We further show the enlargement of interferograms from ALOS-2 and Sentinel-1 data over one of the active mining areas in the NECIB (Figure 10). This mining area is located in the Meihuajing coal mine. Phase fringes due to mining subsidence can be identified by both L-band and C-band DInSAR interferograms, but in the areas with the largest subsidence magnitude, the fringes are difficult to interpret due to phase decorrelation and phase aliasing effects. Therefore, precise estimation of the maximum subsidence magnitude is not attempted, but in many of the interferograms the subsidence

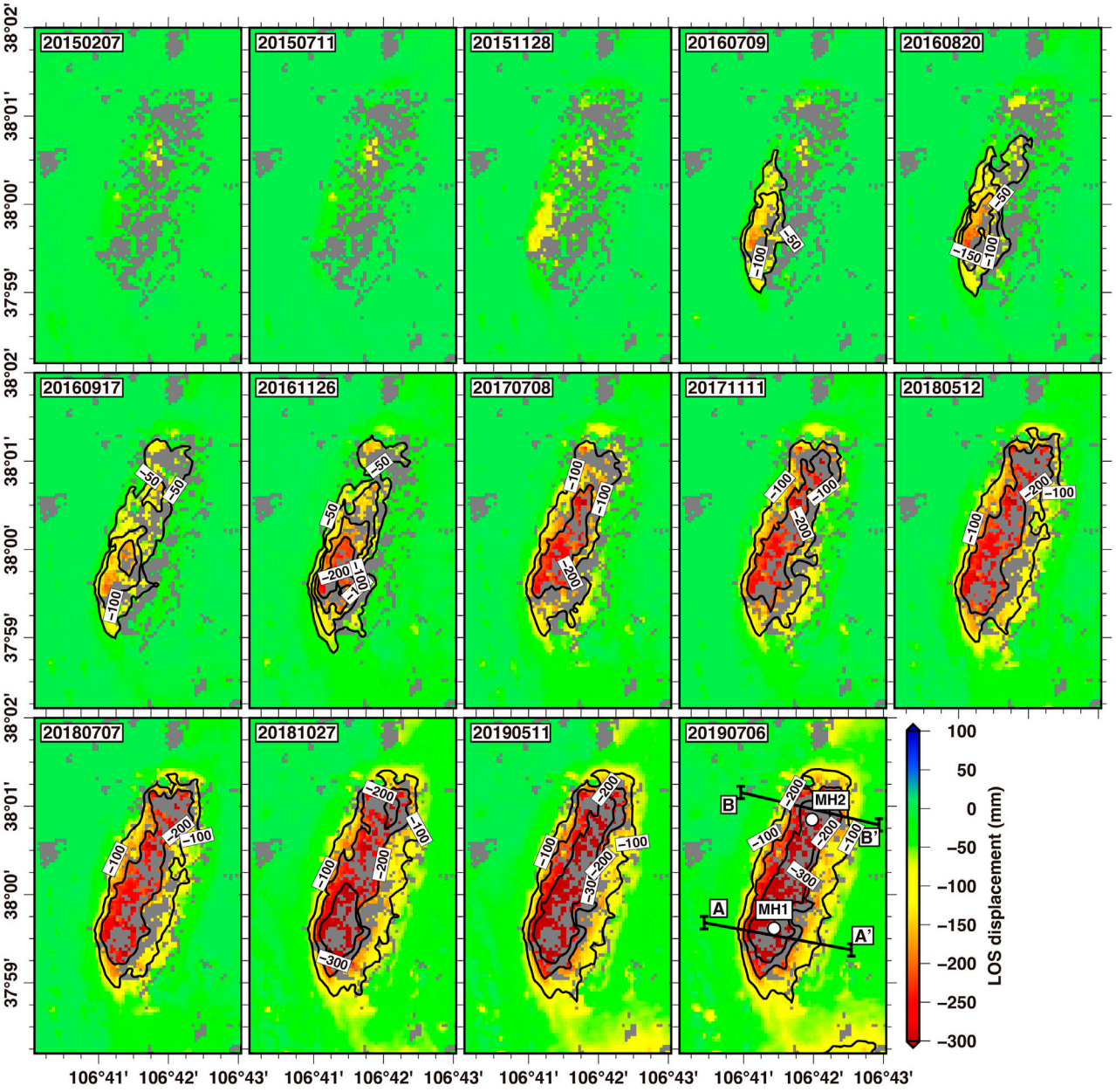


FIGURE 12
The progressive accumulation LOS subsidence over Meihuajing mining area from ALOS-2 data. The black lines on 20190706 (A-A′ and B-B′) are the profiles across the mining subsidence basin, which are shown in Figure 13. The white points (MH1 and MH2) on 20190706 marks the location of the displacement time-series shown in Figure 13.

magnitudes can be reasonably estimated. We can observe a clear increasing number of fringes in the ALOS-2 data with the increase of subsidence magnitude, even on the 20170708 interferogram with a time separation of 588 days. However, the fringes became totally blurred on the last Sentinel-1 interferogram. These phenomena highlight a major limitation of the traditional DInSAR technique, which is not suitable to precisely measure large deformation with significant spatial gradients, because the phase aliasing effects prevent reliable unwrapping. The overall phase quality is low; for ALOS-2 data, the average coherence ranges from a high of 0.34 for the first interferogram (224 days interval) to a low of 0.22 for the last interferogram (588 days interval). For the C-band data of Sentinel-1, although the time gaps between interferometric pairs are shorter than the ALOS-2 data, the coherence loss is much more significant than the L-band data of ALOS-2; the average coherence ranges from a high of 0.23 for the first interferogram (only 24 days interval) to a low of 0.09 for the last interferogram (588 days interval). Except changes in the surface characteristics, very large deformation gradients can cause coherence loss, which can be seen in the mining subsidence center. The effectiveness of DInSAR for

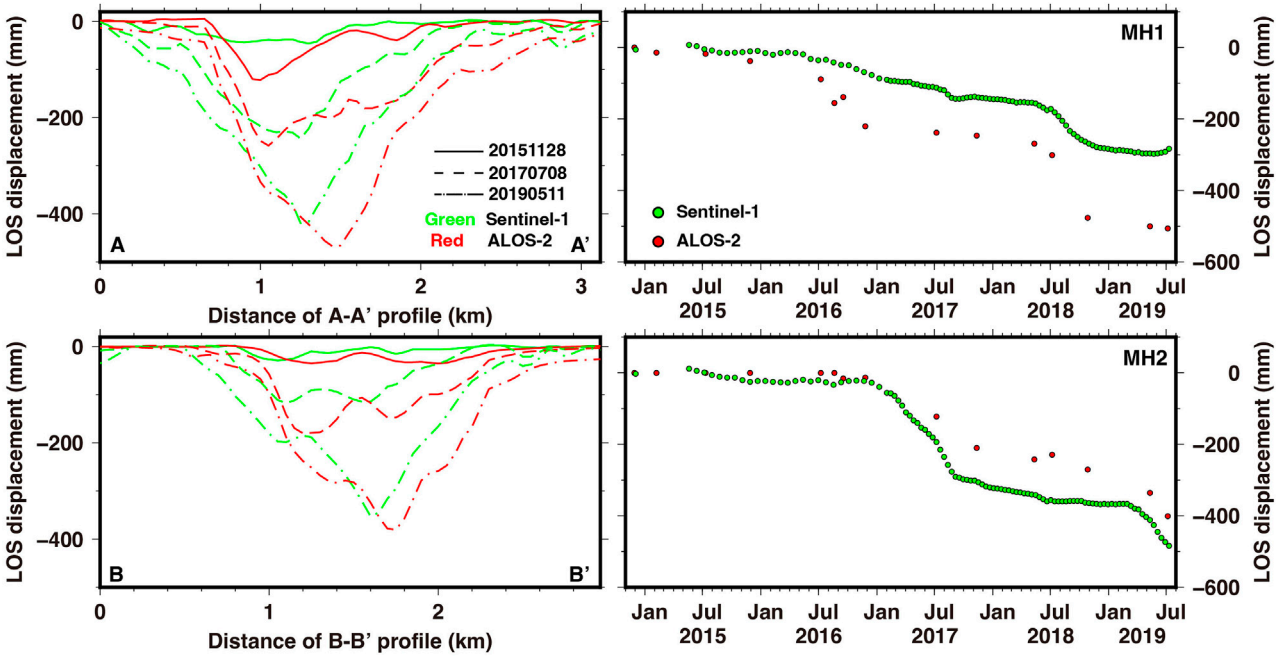


FIGURE 13
Progressively cumulated LOS displacement along profiles A-A′ and B-B′ over Meihuajing mining area from Sentinel-1 and ALOS-2 data (left panel). The LOS displacement time-series at point MP1 and MP2 (right panel). Refer to Figures 11, 12 for the locations of the profiles and points.

quantifying displacement is dependent on the radar wavelength. Generally, subsidence magnitudes are measurable in the ALOS-2 data with L radar band, the Sentinel-1 data with C radar band are more affected by temporal decorrelation and phase aliasing effects, and precisely measuring the subsidence magnitudes using C-band data is more difficult.

5.2 Comparison of cumulated subsidence between Sentinel-1 and ALOS-2

In this section, we compare the cumulated subsidence between Sentinel-1 and ALOS-2 data derived by coherence-based SBAS method to further demonstrate the InSAR capability in resolving the large gradient deformation over the active mining areas. In Figures 11, 12, we show the spatial and temporal evolution of subsidence from Sentinel-1 and ALOS-2 data, respectively, over Meihuajing coal mine in the NECIB. Longwall advance is from northeast to southwest. In the figure, we only displayed the cumulated subsidence on the 14 acquisition dates from Sentinel-1 closest to ALOS-2 acquisition dates in order to compare their results. The cumulated subsidence maps from ALOS-2 and Sentinel-1 data were temporally referenced to a common acquisition date—the first acquisition date on 29 November 2014.

Generally, from Figures 11, 12, we observed very similar spatial extent and magnitude of the cumulated subsidence between Sentinel-1 and ALOS-2 results, although the acquisition dates of the two datasets were not exactly the same. In Figure 13, we further show the cumulated subsidence along two profiles (A-A′ and B-B') on three same acquisition dates (20151128, 20170708 and 20190511) for both Sentinel-1 and ALOS-2 data. And the displacement time-series at two points (MH1 and MH2) over this mine are also plotted in Figure 13. Although the general displacement patterns were consistent along the two profiles, discrepancies between the two SAR sensors can be found, especially during active mining phases (on 20170708 and 20190511) with large spatial displacement gradients, where ALOS-2 could extract a larger magnitude of subsidence than Sentinel-1. From the displacement time-series at MP1 and MP2 in Figure 13, we observed that the subsidence evolution from Sentinel-1 is much smoother than that of ALOS-2. We clearly observed two accelerating patterns in the time series at both MH1 and MH2. From the ALOS-2 displacement time series at MH1, we identified the first accelerating phase was between July 2016 and January 2017 and the second was between July 2018 and November 2018, but Sentinel-1 results did not show the accelerating pattern during the first phase. The subsidence magnitude derived from Sentinel-1 is much smaller than that from ALOS-2. It is probably due to the unreliable unwrapping on those interferograms with large displacement gradient during the active mining phase. At MH2, the first accelerating phase was between January 2017 and August 2017 and the second accelerating phase was between March 2019 and July 2019, both the two results can clearly identify the two accelerating patterns. However, the subsidence magnitude from ALOS-2 is slightly less than that of Sentinel-1. From the displacement time-series plots, we can conclude that the Sentinel-1 acquisition periods are shorter than ALOS-2

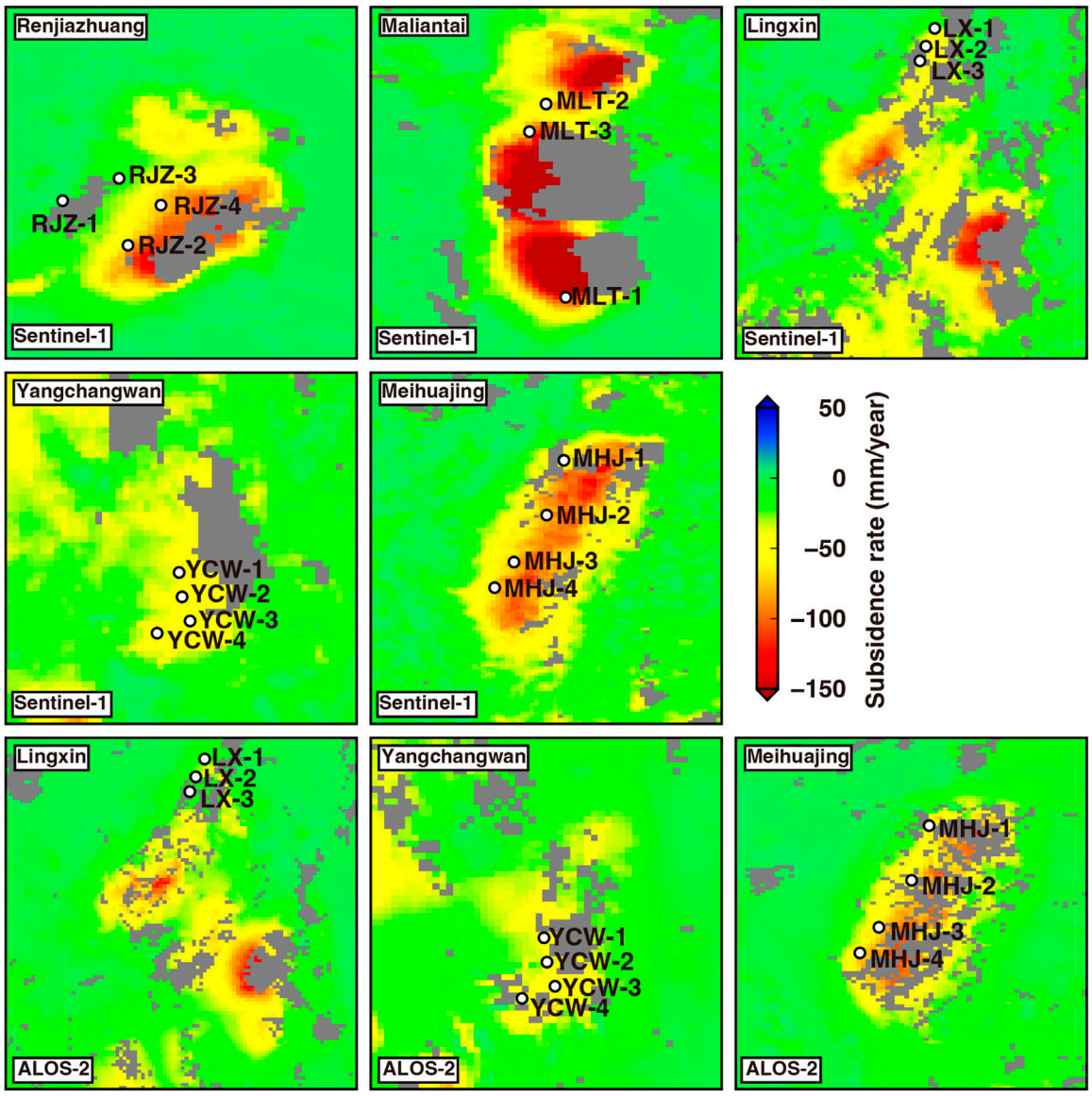


FIGURE 14
The available continuous GPS stations (white circles) overlay on the subsidence rate maps over five mines: Renjiazhuang (RJZ-1, RJZ-2, RJZ-3 and RJZ-4), Maliantai (MLT-1, MLT-2 and MLT-3), Lingxin (LX-1, LX-2 and LX-3), Yangchangwan (YCW-1, YCW-2, YCW-3 and YCW-4) and Meihuajing (MHJ-1, MHJ-2, MHJ-3 and MHJ-4). No ALOS-2 data coverage over Renjiazhuang and Maliantai mines.

acquisition periods, the Sentinel-1 data provide a more timely report of the subsidence development.

5.3 Comparison of subsidence between InSAR and GPS

In this section, we compare InSAR displacement time series with that of GPS to validate the accuracy of InSAR-derived subsidence. A total of 18 continuous GPS monitoring stations located in five mining areas (Renjiazhuang, Maliantai, Lingxin, Yangchangwan and Meihuajing mines) are available for the comparison (Table 3). The locations of GPS stations are superimposed on the subsidence rate maps, as shown in Figure 14. The comparison of displacement time series between these two techniques is shown in Figure 15. GPS can only obtain displacement at sparse discrete points. To match the InSAR pixels and the continuous GPS stations, we averaged InSAR pixels within a circular area with a radius of 50 m centered on the GPS station for the comparison. The 3D NEU displacement resulting from GPS analysis were first projected into the InSAR LOS direction as:

$$d_{LOS} = \begin{bmatrix} \sin\theta\sin\alpha & -\sin\theta\cos\alpha & \cos\theta \end{bmatrix} \begin{bmatrix} d_N \\ d_E \\ d_U \end{bmatrix} \tag{8}$$

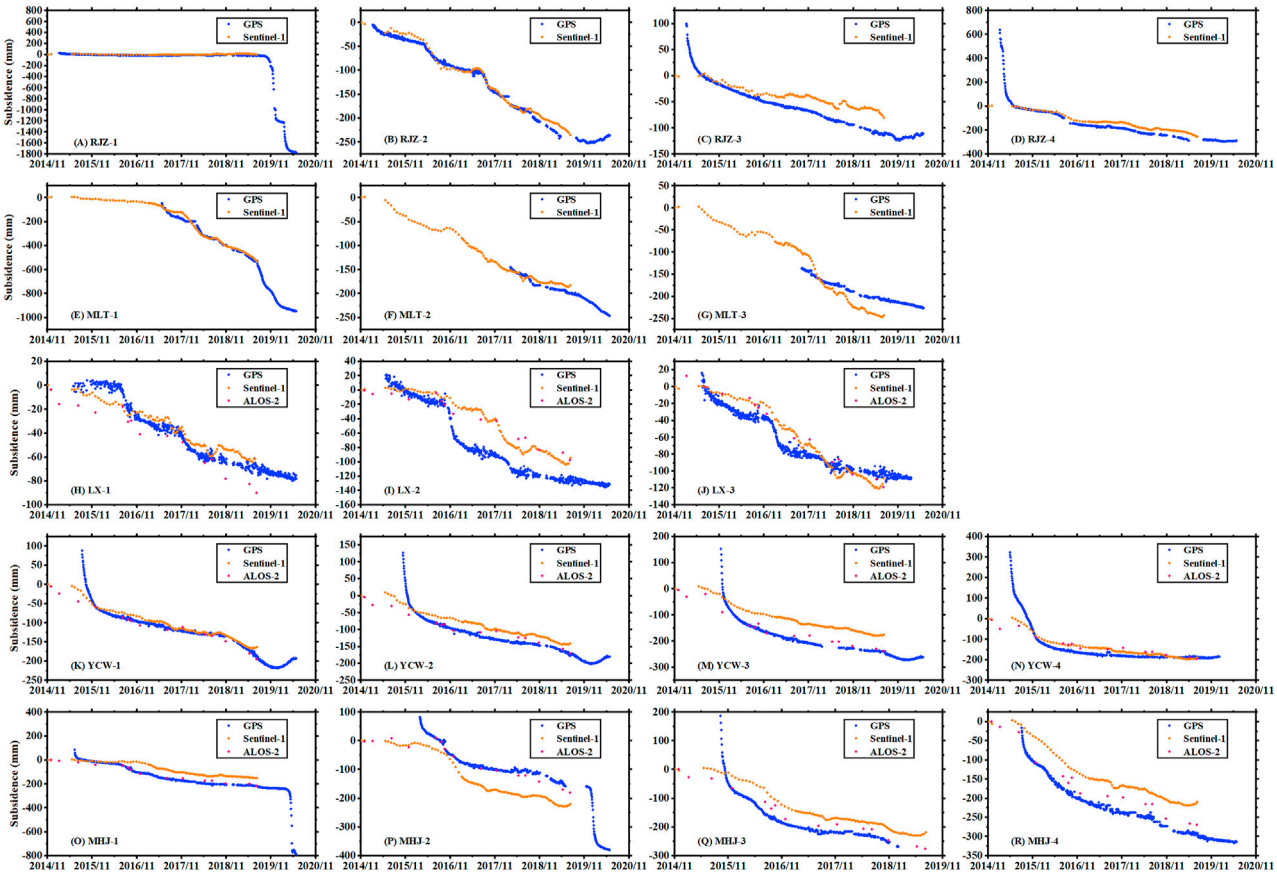


FIGURE 15
Comparison between InSAR- and GPS-derived subsidence time series at five coal mines **(A–D)** Renjiazhuang (RJZ-1, RJZ-2, RJZ-3 and RJZ-4) **(E–G)** Maliantai (MLT-1, MLT-2, and MLT-3) **(H–J)** Lingxin (LX-1, LX-2 and LX-3) **(K–N)** Yangchangwan (YCW-1, YCW-2, YCW-3 and YCW-4) **(O–R)** Meihuajing (MHJ-1, MHJ-2, MHJ-3 and MHJ-4). The blue dots indicate GPS observation, the orange dots represent the Sentinel-1 observations, and red dots are ALOS-2 measurements. No ALOS-2 data coverage over Renjiazhuang and Maliantai mines.

where α is the heading angle of the SAR satellite (the satellite's orbit angle relative to true north), θ is the incidence angle of the radar wave.

In Figure 15, over the Renjiazhuang mine, a very good agreement between the observations between InSAR and GPS can be found at RJZ-1 station (Figure 15A) because the ground at this location was stable during the InSAR time period. This area had an abrupt subsidence in November 2019 as observed from the GPS measurements. We observed a good agreement between the observations between InSAR and GPS at RJZ-1, RJZ-2 and RJZ-4. A good agreement between these two techniques was also found at RJZ-2 station (Figure 15B) because the displacement is not significant in terms of mining subsidence. However, we found an abrupt change between November 2014 and May 2015 from GPS observations at RJZ-3 and RJZ-4 stations (Figures 15C, D) where InSAR measurements cannot capture such large displacement gradient due to phase aliasing effects. These phenomena can also be found at GPS stations over Yangchangwan and Meihuajing mines (Figures 15K–R). Generally, good agreements between InSAR and GPS can be observed if the displacements are continuous. We further show the scatter plot between InSAR and GPS subsidence measurements in Figure 16. For Sentinel-1 data, the average RMSE between InSAR and GPS measurements is 28.4 mm, and for each mine the RMSE is 24.1, 22.9, 15.3, 20.1, 59.5 mm over Renjiazhuang, Maliantai, Lingxin, Yangchangwan, and Meihuajing mine, respectively. For ALOS-2 data, the average RMSE between InSAR and GPS measurements is 21 mm, and for each mine the RMSE is 21.6, 15.8, 25.6 mm at Lingxin, Yangchangwan, and Meihuajing mine, respectively. When comparing the Sentinel-1 and ALOS-2 data over Lingxin mine (Figures 15H–J), a high consistency among Sentinel-1, ALOS-2 and GPS measurements was observed because the stations located in areas experiencing mild deformation. However, at YCW-3 (Figure 15M), MHJ-3 (Figure 15Q) and MHJ-4 (Figure 15R), the ALOS-2 data are closer to the GPS measurements because the significant deformation over these stations that ALOS-2 data has a better performance at capturing rapid subsidence.

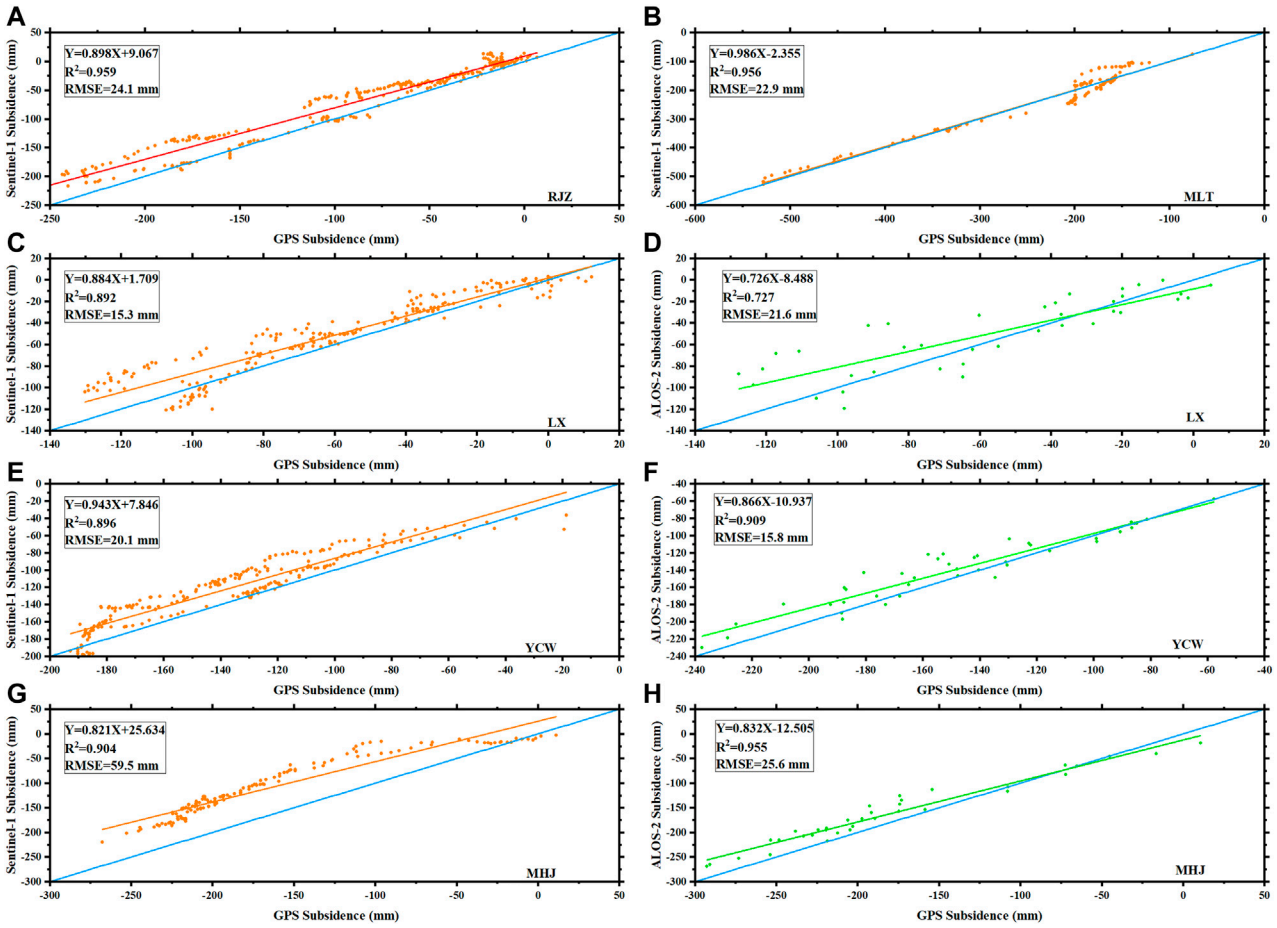


FIGURE 16
The scatter plot between GNSS- and InSAR-derived subsidence.

6 Conclusion

Differential Synthetic Aperture Radar Interferometry has been used to detect and monitor fast mining-induced subsidence over Ningdong coal base area located in northwestern China. A total of 15 L-band SAR images from ALOS-2 satellite and 102 C-band images from Sentinel-1 satellite spanning between November 2014 and July 2019 have been processed using both stacking and SBAS time series analysis techniques. At a regional scale over a 3,484 km^2 area of the coal base, InSAR-derived subsidence maps have been compared with the boundary of ten mining concessions to identify those mining areas generating significant subsidence and to delimit the extent of the area affected by mining subsidence. At a local scale in all the active coal mines, the spatial and temporal evolution of subsidence were obtained to track the subsidence dynamics and determine the characteristic surface response to longwall advance. The most significant subsidence in the study region was found at Maliantai mine with −264 mm/year detected by SBAS method from Sentinel-1 data.

The effectiveness of DInSAR for estimating mining subsidence is dependent on the radar wavelength. Compared to C-band, L-band SAR are less affected by phase aliasing due to large deformation gradients and by temporal decorrelation due to changes in surface scattering characteristics over time. With a shorter imaging period (12 days), the Sentinel-1 data improve the capability in quantifying the fast mining subsidence and provide a more timely report of the subsidence extent.

We validated the InSAR displacement time-series by comparing both ALOS-2 and Sentinel-1 results with 18 continuous GPS stations above five active mining regions in this coal base. When comparing the displacement time series between InSAR and GPS, we found that InSAR time series method cannot capture the abrupt mining subsidence with large displacement gradient due to phase aliasing effects. The GPS measurements with a high sampling rate in time can properly observe such abrupt change. We can combine these two techniques to obtain the full three-dimensional, large displacements due to mining activities in this coal base area.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

WT wrote the manuscript, designed its structure, and analyzed the data collected. MW helped in collecting and analyzing the data. PL and GW provided access to the data, support in reviewing and editing the manuscript. YY and WY critically revised the manuscript. All authors contributed to the article and approved the submitted version..

Funding

This work is supported by Open Fund of State Key Laboratory of Water Resource Protection and Utilization in Coal Mining (Grant No. WPUKFJJ 2019-17), Open Fund of State Key Laboratory of Coal Resources and Safe Mining (Grant No. SKLCRSM22KFA08), Open Fund of Key Laboratory of Natural Resources Monitoring and Supervision in Southern Hilly Region, Ministry of Natural Resources (Grant No. NRMSSHR 2022Y05). The authors declare that this study received funding from State Key Laboratory of Water Resource Protection and Utilization in Coal Mining, CHN Energy Shendong Coal Group Co., Ltd., Beijing, China, National Institute of Clean-and-Low-Carbon Energy, Beijing, China, and Key Laboratory of Natural Resources Monitoring and Supervision in Southern Hilly Region, Ministry of Natural Resources, Changsha, China. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

The authors would like to thank the European Space Agency for providing Sentinel-1 data and the Alaska Satellite Facility for downloading the archived Sentinel-1 data. The authors thank the Japanese Space Exploration Agency (JAXA) for providing ALOS-2 data. The authors would also like to thank the Institute of Land and Resources Investigation and Monitoring, Ningxia Hui Autonomous Region, for technical support during field investigation. The authors are grateful to the reviewers for their constructive comments.

Conflict of interest

WT was employed by CHN Energy Shendong Coal Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Berardino, P., Fornaro, G., Lanari, R., and Sansosti, E. (2002). A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans. Geoscience Remote Sens.* 40, 803792. doi:10.1109/TGRS.2002.803792

Cai, M., Wu, Y., Li, P., Pan, J., Hong, W., and Li, J. (2022). Present situation and ideas of green development of coal resources in Ningxia. *Chin. J. Eng.* 44, 1–10.

Chen, C. W., and Zebker, H. A. (2001). Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization. *J. Opt. Soc. Am. A* 18, 338–351. doi:10.1364/josaa.18.000338

Chen, J., Ban, Y., and Li, S. (2014). China: Open access to Earth land-cover map. *Nature* 514, 434. doi:10.1038/514434c

Dai, Y., Ng, A. H. M., Wang, H., Li, L., Ge, L., and Tao, T. (2021). Modeling-assisted InSAR phase-unwrapping method for mapping mine subsidence. *IEEE Geoscience Remote Sens. Lett.* 18, 1059–1063. doi:10.1109/lgrs.2020.2991687

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Rev. Geophys.* 45, 1–33. doi:10.1029/2005rg000183

Ferretti, A., Prati, C., and Rocca, F. (2001). Permanent scatterers in SAR interferometry. *IEEE Trans. Geoscience Remote Sens.* 39, 8–20. doi:10.1109/36.898661

Hanssen, R. (2001). *Radar Interferometry: Data interpretation and error analysis*. New York: Kluwer Academic Publishers. Available at: https://books.google.co.uk/books?hl=en&lr=&id=bqNkJUk4wtMC&oi=fnd&pg=PA4&dq=Radar+Interferometry:+Data+Interpretation+and+Error+Analysis&ots=8NjwnEA1fL&sig=nb9LBEZq7o_lRiB85MPLFRTJyh8.

Hooper, A., Segall, P., and Zebker, H. (2007). Persistent scatterer interferometric synthetic aperture radar for crustal deformation analysis, with application to Volcán Alcedo, Galápagos. *J. Geophys. Res. Solid Earth* 112, B07407–B07421. doi:10.1029/2006jb004763

Ilieva, M., Polanin, P., Borkowski, A., Gruchlik, P., Smolak, K., Kowalski, A., et al. (2019). Mining deformation life cycle in the light of InSAR and deformation models. *Remote Sens.* 11, 745. doi:10.3390/rs11070745

Liu, J., Liu, X., Lv, X., Wang, B., and Lian, X. (2022). Novel method for monitoring mining subsidence featuring Co-registration of UAV LiDAR data and photogrammetry. *Appl. Sci.* 12, 9374. doi:10.3390/app12189374

Modeste, G., Doubre, C., and Masson, F. (2021). Time evolution of mining-related residual subsidence monitored over a 24-year period using InSAR in southern Alsace, France. *Int. J. Appl. Earth Observation Geoinformation* 102, 102392. doi:10.1016/j.jag.2021.102392

Przylucka, M., Herrera, G., Graniczny, M., Colombo, D., and Bejar-Pizarro, M. (2015). Combination of conventional and advanced DInSAR to monitor very fast mining subsidence with TerraSAR-X data: Bytom City (Poland). *Remote Sens.* 7, 5300–5328. doi:10.3390/rs70505300

Sandwell, D. T., and Price, E. J. (1998). Phase gradient approach to stacking interferograms. *J. Geophys. Res. Solid Earth* 103, 30183–30204. doi:10.1029/1998JB900008

Schmidt, D. A., and Bürgmann, R. (2003). Time-dependent land uplift and subsidence in the Santa Clara valley, California, from a large interferometric synthetic aperture radar data set. *J. Geophys. Res. Solid Earth* 108. doi:10.1029/2002jb002267

Smith, W. H. F., and Wessel, P. (2012). Gridding with continuous curvature splines in tension. *Geophysics* 55, 293–305. doi:10.1190/1.1442837

Tang, W., Zhao, X., Motagh, M., Bi, G., Li, J., Chen, M., et al. (2022). Land subsidence and rebound in the Taiyuan basin, northern China, in the context of inter-basin water transfer and groundwater management. *Remote Sens. Environ.* 269, 112792. doi:10.1016/j.rse.2021.112792

Tong, X., and Schmidt, D. (2016). Active movement of the Cascade landslide complex in Washington from a coherence-based InSAR time series method. *Remote Sens. Environ.* 186, 405–415. doi:10.1016/j.rse.2016.09.008

Tymofyeyeva, E., and Fialko, Y. (2015). Mitigation of atmospheric phase delays in InSAR data, with application to the eastern California shear zone. *J. Geophys. Res. Solid Earth* 120, 5952–5963. doi:10.1002/2015jb011886

Wang, G., Xu, Y., Cheng, X., and Zhang, J. (2018). A study of GPS ground deformation monitoring of Jinfeng coal mine in Ningdong coal base. *Geol. Bull. China* 37, 2199–2207.

Wang, G., Wu, Q., Li, P., Cui, X., Gong, Y., Zhang, J., et al. (2021). Mining subsidence prediction parameter inversion by combining GNSS and DInSAR

deformation measurements. *IEEE Access* 9, 89043–89054. doi:10.1109/access.2021.3089820

Wegmuller, U., Strozzi, T., Werner, C., Wiesmann, A., Benecke, N., and Spreckels, V. (2000). Monitoring of mining-induced surface deformation in the Ruhrgebiet (Germany) with SAR interferometry. *Int. Geoscience Remote Sens. Symposium* 6, 2771–2773. doi:10.1109/IGARSS.2000.859710

Wempen, J. M., and McCarter, M. K. (2017). Comparison of L-band and X-band differential interferometric synthetic aperture radar for mine subsidence monitoring in central Utah. *Int. J. Min. Sci. Technol.* 27, 159–163. doi:10.1016/j.ijmst.2016.11.012

Wempen, J. M. (2020). Application of DInSAR for short period monitoring of initial subsidence due to longwall mining in the mountain west United States. *Int. J. Min. Sci. Technol.* 30, 33–37. doi:10.1016/j.ijmst.2019.12.011

Wessel, P., Luis, J., Uieda, L., Scharroo, R., Wobbe, F., Smith, W., et al. (2019). The generic mapping Tools version 6. *Geochem. Geophys. Geosystems* 20, 5556. doi:10.1029/2019GC008515

Wu, Z., Wang, T., Wang, Y., Wang, R., and Ge, D. (2022). Deep learning for the detection and phase unwrapping of mining-induced deformation in large-scale interferograms. *IEEE Trans. Geoscience Remote Sens.* 60, 1–18. doi:10.1109/tgrs.2021.3121907

Xu, X., Sandwell, D. T., Tymofyeyeva, E., Gonzalez-Ortega, A., and Tong, X. (2017). Tectonic and anthropogenic deformation at the cerro prieto geothermal step-over revealed by sentinel-1A InSAR. *IEEE Trans. Geoscience Remote Sens.* 55, 5284–5292. doi:10.1109/tgrs.2017.2704593

Zhang, L., Dai, K., Deng, J., Ge, D., Liang, R., Li, W., et al. (2021). Identifying potential landslides by stacking-InSAR in southwestern China and its performance comparison with SBAS-InSAR. *Remote Sens.* 13, 3662. doi:10.3390/rs13183662

Zhu, C., Wu, W., Motagh, M., Zhang, L., Jiang, Z., and Long, S. (2020). Assessments of land subsidence along the Rizhao-Lankao high-speed railway at Heze, China, between 2015 and 2019 with Sentinel-1 data. *Nat. Hazards Earth Syst. Sci.* 20, 3399–3411. doi:10.5194/nhess-20-3399-2020

Frontiers in Earth Science

Investigates the processes operating within the major spheres of our planet

Advances our understanding across the earth sciences, providing a theoretical background for better use of our planet's resources and equipping us to face major environmental challenges.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact