Transportation Research Part D 136 (2024) 104422

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Assessment and modeling of roadside geological risks in the Qinghai-Tibetan Plateau region

Hong Zhang[a], Xin Xu[b,*], Chi Zhang[c,d,**], Hong-Zhi Yang[c], Yiik Diew Wong[e]

[a] *Transportation Institute, Inner Mongolia University, Hohhot 010020, China*
[b] *Pioneer College, Inner Mongolia University, Hohhot 010020, China*
[c] *Key Laboratory for Special Area Highway Engineering of Ministry of Education, Chang'an University, Xi'an 710064, Shaanxi, China*
[d] *Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an 710000, Shaanxi, China*
[e] *School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore*

ARTICLE INFO

Keywords:
Qinghai-Tibetan Plateau
Mountain geological hazards
Frozen soil thermal stability
Hotspot analysis
Normalized spectral entropy
Generalized Estimation Equations

ABSTRACT

This study identified and assessed spatiotemporal distribution, and modeled influencing factors, of roadside geological risks for the Qinghai-Tibetan Plateau regions towards enhancing the safety of road operations. It utilized Latent Dirichlet Allocation model to identify risks, hotspot analysis and normalized spectral entropy to assess the spatiotemporal risk distribution, and Generalized Estimation Equations to analyze the relationship between geological risk susceptibility and the influencing factors. Finally, a case study was conducted to apply this methodology. The findings indicated that mountain geological hazards could affect areas up to 2,000 m from the roadway. Roadbed and tunnel sections, and lower-grade highways were more susceptible to mountain geological hazards, while higher-grade highways tended to have lower frozen soil thermal stability in their roadbed sections. This study provides valuable insights into the coupling effects between engineering and geological environments, crucial for informed route layout decisions and effective management of roadside geological risks.

1. Introduction

The Qinghai-Tibetan Plateau (QTP) is known for its uplifted plateau and intensive tectonic activities, characterized by numerous mountains, glacial lakes, significant terrain variations, and complex geological conditions (Zeng et al., 2018; Zhang et al., 2019). These conditions contribute to frequent occurrences of various geohazards such as creep slopes, landslides, collapses, and active faults (Wang et al., 2020; Xia et al., 2021). Additionally, the presence of groundwater coupled with cold climate leads to frozen soil distresses such as ice cones, frost heave mounds, and thermokarst landslides (Du et al., 2021; Zhang et al., 2019). Numerous studies have been conducted in the QTP regions to better understand these phenomena. For instance, Jia et al. (2021) analyzed the spatial transfer patterns of slope geohazards and their correlation with temperature and precipitation. Wang et al. (2020) assessed the risk of glacier lake outburst floods. Ni et al. (2022) explored the relationship between thawing hazards and human activities, including land use, road distribution, population density, and grazing intensity. Yin et al. (2021) employed machine learning models to investigate the causes of thermokarst

---

* Corresponding author at: Xilinguole South Road, Yuquan District, Hohhot 010020, Inner Mongolia, China.
** Corresponding author at: Middle Section of the South Second Ring Road, Yanta District, Xi'an 710064, Shaanxi, China.
E-mail addresses: hongzhang@chd.edu.cn (H. Zhang), xuxin@chd.edu.cn (X. Xu), zhangchi@chd.edu.cn (C. Zhang), yhz@chd.edu.cn (H.-Z. Yang), CYDWONG@ntu.edu.sg (Y. Diew Wong).

https://doi.org/10.1016/j.trd.2024.104422
Received 5 December 2023; Received in revised form 4 August 2024; Accepted 10 September 2024

landslides. Jiao et al. (2023) utilized radar data to map the spatiotemporal patterns of ground deformation caused by retrogressive thaw slumps. These geohazards pose significant challenges to the design, construction, and operation of highways and railways, greatly impacting traffic safety and substantially increasing construction and maintenance costs (Cui et al., 2022).

Planners, builders, operators, and scholars in the field of transportation infrastructure consistently prioritize geosafety issues, particularly in mountainous regions (Solorzano et al., 2022). Recent research into coupling of transportation and geology has primarily focused on identifying geohazards, assessing their susceptibility and influencing factors (Amatya et al., 2019; He et al., 2021; Liang et al., 2021; Liu et al., 2023; Zhang et al., 2020), as well as studying slope instability and engineering protective measures (Li et al., 2023; Lu et al., 2013). While recent achievements have shown promising practical results, there remains limited quantitative and mechanistic analyses regarding the relationship between the susceptibility of geological risks and the attributes of highways or railways, along with relevant natural environmental factors. Therefore, the objective of this study is to assess the spatiotemporal distribution of geological risks along roadsides, identify high-risk locations, extract both road attributes and natural factors of these locations for modeling and analysis, and investigate the coupling effects between route engineering and geological environments. The contributions of this study to the literature are reflected in proposing a comprehensive method for assessing and modeling roadside geological risks. The findings hold significant importance for the planning of highway and railway corridors, as well as for mitigating roadside geological risks in the QTP regions. Moreover, the methodology and findings are also applicable to other similar areas with sensitive and fragile geologic environments.

In terms of identifying geological risks, commonly used methods include field surveys, expert interviews, qualitative analysis, remote sensing image analysis, radar measurements, and machine learning techniques (Abdulla et al., 2018; Luo et al., 2021; Wang et al., 2023; Zhao et al., 2022). While field surveys are reliable, they are limited in scope, time-consuming, and costly. Expert interviews and qualitative analysis can effectively determine the types of disasters, but they often lack objectivity. Other methods that utilize satellite image data to identify the location, quantity, and scale of disasters require high computational power when applied to large-scale areas. These methods are aimed at identifying single geohazards within designated areas where data acquisition is relatively easy. In new research areas with scarce data and unclear geological conditions, a reliable approach for determining the types of disasters is through keyphrase extraction, a text mining technique that analyzes geological research hotspots in a specific region. This

Table 1
Comparison of common text mining techniques for keyphrase extraction.

Text mining techniques	Principles	Advantages	Disadvantages	Specific algorithms	References
Statistical features-based methods	▪ Extracting keywords from documents based on metrics such as word frequency, lexicality, and position.	▪ Simple in principle and easy to model.	▪ Disregarding the distribution bias of words in different documents; ▪ Requiring a high-quality corpus that matches the text being processed for training; ▪ Resulting in lower accuracy.	TF, TF-IDF, YAKE	Campos et al. (2018); Justicia De La Torre et al. (2018)
Word graph models-based methods	▪ Constructing co-occurrence, syntactic, and semantic networks, as well as other document graphs, to identify keywords through their inter-connected relationships.	▪ Evaluating the significance of various words from a holistic perspective.	▪ Relying on language network graphs and contextual relationships; ▪ Exhibiting high computational complexity and having restricted applicability.	PageRank, TextRank	Mihalcea and Tarau (2004)
Topic models-based methods	▪ Identifying words highly relevant to a topic by analyzing their distribution within that specific context.	▪ Considering the distribution of topics and keywords across different documents.	▪ Extracted keywords may be too general to effectively represent the document's theme; ▪ Requiring substantial practical training to implement.	LDA, pLSA, MDERank, NMF	Zhang et al. (2022); Jelodar et al. (2019); Onan (2018)
Supervised methods	▪ Employing classification, regression, or other machine learning algorithms on annotated data to identify keywords.	▪ Offering higher predictive accuracy; ▪ Integrating multiple features and algorithms to adapt to various text analysis needs.	▪ Requiring prior labeling of the training set, leading to high manual costs.	RAKE, keyBERT, word2vec, GenEx, TextBlob, Bayesian, SVM, GBDT, LR-SGD	Rose et al. (2010); Tandel et al. (2019)

approach overcomes subjective issues and eliminates the need for high-performance computers. Common text mining techniques for keyphrase extraction include unsupervised methods based on statistical features, word graph models, and as well as principles, advantages, disadvantages, and specific algorithms are outlined in Table 1. Among these, the topic-based Latent Dirichlet Allocation (LDA) model employs unsupervised learning to directly obtain the topic classifications and probability distributions of keywords in documents (Jelodar et al., 2019; Onan, 2018). The algorithm is simple, and its results are easily interpretable and understandable, yielding better outcomes in hotspot research analysis. Therefore, the LDA model was employed to identify themes in literature related to geohazards in the QTP and to summarize the associated geological risks.

Apart from identification of geohazards, scholars have extensively studied the assessment of geohazard susceptibility. These studies have employed various methods, including mathematical model-driven approaches (Chang et al., 2021; Sahoo et al., 2016; Tan et al., 2021), data-driven techniques (Dai et al., 2020; Gao et al., 2022; Niu, 2020; Yang et al., 2019; Yu et al., 2017), and spatial modeling approaches that are typically conducted based on Geographic Information System (GIS) technologies (Ambrosi et al., 2018; Kusumawati et al., 2017; Ourang et al., 2019; Qi et al., 2015; Zhang et al., 2021). Specific examples of these approaches are illustrated in Fig. 1. For instance, Lin et al. (2019) established a risk assessment indication system for highway slope geohazards and implemented uncertainty assessment using index importance ranking, thereby reducing subjectivity in the rating process. Zou and Gui (2020) utilized time series analysis to study geohazard data and developed a geohazard prediction model by employing a computer simulation system and building information modeling (BIM). Zhang et al. (2021) utilized remote sensing images and hotspot analysis to conduct spatial clustering of landslide deformations for landslide identification and deformation analysis. These mathematical model-driven approaches have been widely applied and are considered quite mature, but a rational and scientific evaluation indication system is still lacking. Data-driven approaches provide flexibility and convenience in modeling but rely on extensive data, with many methods still under development. Spatial modeling approaches offer more intuitive and practical solutions but lack quantitative analysis in the relationship between geological risk susceptibility and its influencing factors. To address these gaps, the present study integrated the advantages of spatial modeling for visualization with mathematical model-driven approaches for quantitative analysis. Specifically, GIS-based hotspot analysis was employed to evaluate the spatial distribution of past geohazards. This method can identify areas with statistically significant clustering in space, and is suitable for analyzing large datasets. Subsequently, Generalized Estimation Equations (GEEs) were established to analyze the relationship between geohazard susceptibility and various influencing factors. GEEs are well-suited for modeling geographic data that follow different distributions, and can analyze non-independent categorical and continuous variables in repeated measurements, such as geohazards and ground temperature (Onder, 2016; Zhang et al., 2020).

When examining geologic risk associated with frozen soil, traditional studies have primarily focused on the impacts of climate change and permafrost on roadbed settlement and stability (Fortier et al., 2011; Ghias et al., 2017). Several studies have also explored the coupling effects between temperature fields and thermal-hydrological fields in permafrost regions (Liu et al., 2019). These studies often rely on field measurements of ground temperature or numerical simulation methods to assess permafrost thermal stability (Liu et al., 2019; Makarieva et al., 2018). However, the availability of field measurement data is often limited, and the available point data may not accurately represent the spatial distribution of ground temperature (Li et al., 2019). Furthermore, numerical simulation results rely heavily on model performance. These studies overlook the long-term cumulative impacts of road engineering on frozen soil and fail to fully consider the coupling effect of frozen soil thermal stability and road engineering from a spatiotemporal perspective. Additionally, there is a lack of long-term observations and effective methods for quantitatively evaluating how road engineering

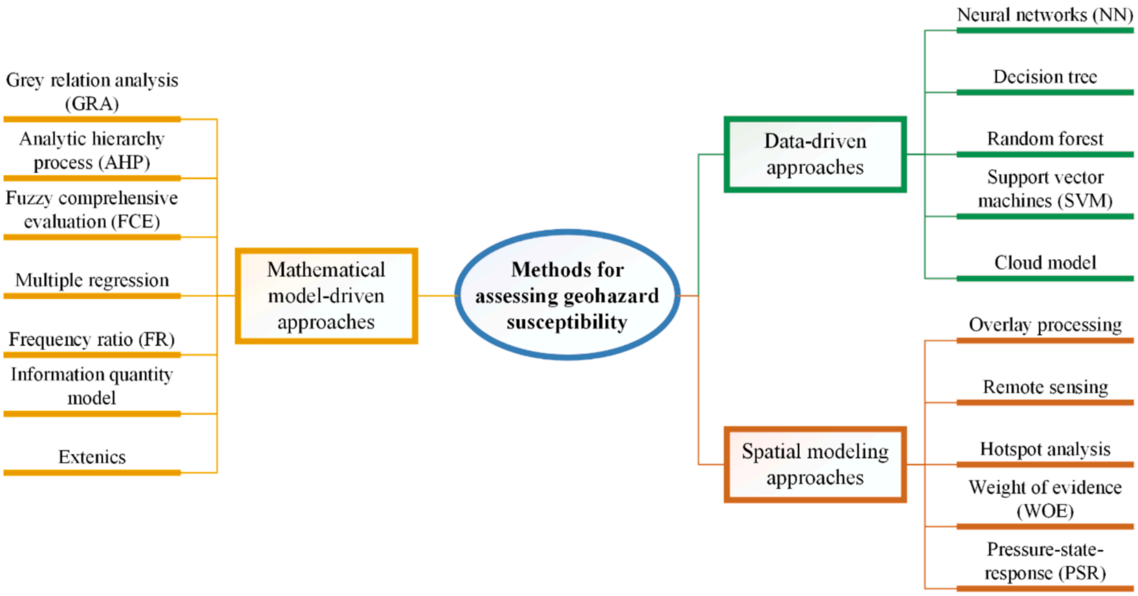

Fig. 1. Methods for assessing geohazard susceptibility.

activities affect frozen soil thermal stability. To address these gaps, this study introduced the normalized spectral entropy to analyze the variation in ground surface temperature from a spatiotemporal perspective, aiming to characterize the thermal stability of frozen soil. The spectral entropy can quantify the complexity and irregularity of time series, providing an effective means for feature extraction from time series data.

All in all, the LDA model was adopted to identify geological risks in the QTP regions. Then, GIS-based hotspot analysis and normalized spectral entropy were employed to evaluate the spatiotemporal distribution of roadside geological risks. GEEs were established to quantitatively analyze the relationship between risk susceptibility and various influencing factors. Finally, a case study was conducted to validate this methodology in the Western Sichuan Plateau region of China, located in the southeast of the QTP. In terms of contributions, previous studies evaluating geohazard susceptibility have primarily been conducted at a regional or slope level, considering factors such as topography, geomorphology, and natural climate. In contrast, the present study focused on roadside areas in the QTP regions, and considered not only traditional environmental factors but also the synergistic effects of road attributes (e.g., road length, structure, grade, and operation duration). This methodology promotes proactive design of transportation infrastructure that enhances the traffic safety and sustainability of highway operations.

2. Research methodology

This study proposed a set of methods for identifying, assessing, and modeling roadside geologic risks, with the modeling process grounded in the evaluation of actual geohazard data. The identification method (i.e., the LDA model) is applicable to regions with limited data and uncertain geological conditions. Evaluation techniques (i.e., GIS-based hotspot analysis and the normalized spectral entropy model) are effective in areas with available historical geohazard data. Furthermore, the modeling outcomes (i.e., GEEs results) offer predictive insights into potential geohazards in roadsides where such events have not yet occurred. This methodology enables simultaneous assessment and prediction of geologic risks, as depicted in the framework shown in Fig. 2.

2.1. Identification of roadside geological risks

The geologic risks found in the QTP regions may differ from those in general mountainous or plain areas, featuring hazards such as permafrost, avalanches, ice tongue, and scree slopes. To identify the specific roadside geological risks in this region, text mining technology was applied. Specifically, the LDA model was used to extract topics and keywords from relevant literature. The results were then compared and validated with the findings from geohazard field surveys conducted in the QTP regions, in order to confirm and determine the roadside geological risks.

The LDA model assumes that documents are composed of K topics, and each topic is comprised of T keywords. It employs Dirichlet distributions with parameters $\vec{\alpha}$ and $\vec{\beta}$ as the conjugate prior distributions for topics and keywords, respectively (Maier et al., 2018). The probability density functions of topics or keywords are represented by equations (1) and (2).

$$p(\vec{x}|\vec{\alpha}) = \mathrm{Dir}(\vec{x}|\vec{\alpha}) = \frac{\Gamma(\widehat{\alpha})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\prod\nolimits_{i=1}^{k} x_i^{\alpha_i-1} \tag{1}$$

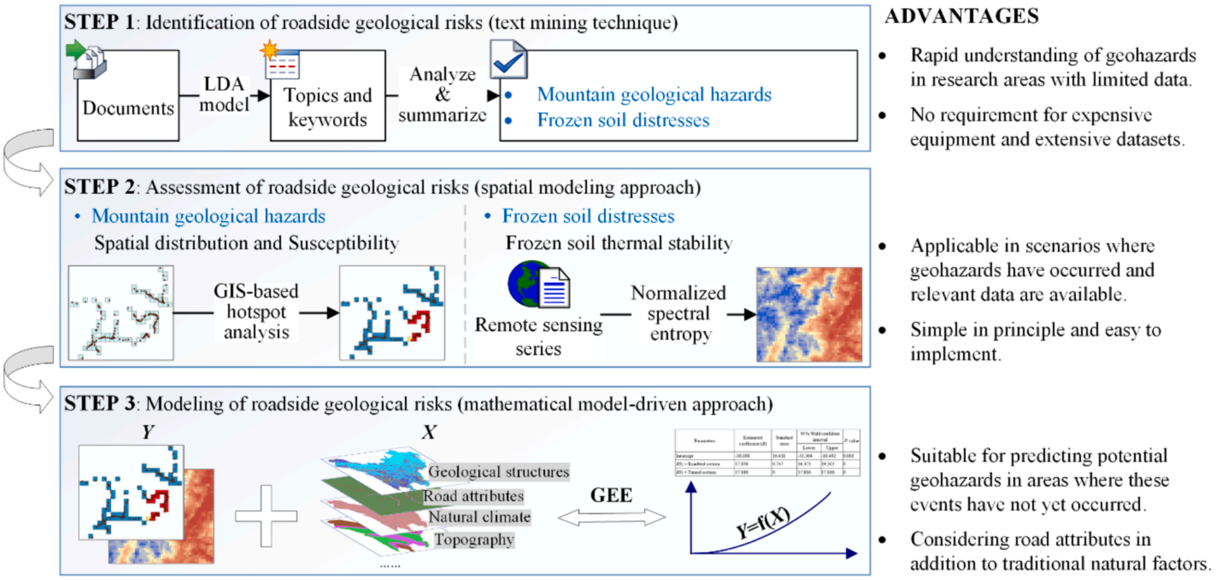

Fig. 2. The methodological framework.

$$\widehat{\alpha} = \sum_{i=1}^{k} \alpha_i \tag{2}$$

The process of identifying roadside geological risks in the QTP regions based on the LDA model is illustrated in Fig. 3, comprising the following steps:

(1) Literature search and stop words compilation. The Chinese Science Citation Database and the Web of Science Core Collection were searched using the terms “Qinghai-Tibet OR High-Altitude AND Geology” from 2015 onwards. Relevant records were selected, and keywords were exported as text files. A total of 489 records (273 in English and 216 in Chinese) were obtained. Chinese stop words were compiled from various sources such as Harbin Institute of Technology University, Sichuan University Machine Intelligence Laboratory, and Baidu, while English stop words were obtained from the NLTK package (Banks et al., 2018).

(2) Model training. The retrieved literature records and stop words served as input data. The Jieba algorithm was used for word tokenization, and the Gensim package in Python was utilized to train the LDA model. Different numbers of topics and keywords were tested, and the output results were compared to minimize redundancy. Ultimately, 3 topics each containing 5 keywords were selected.

(3) Model optimization and result output. The model was optimized by adjusting the stop words and eliminating irrelevant keywords based on the output keywords. The model was then retrained, with the optimized results presented in Table 2.

According to Table 2, the Chinese keywords with higher probabilities (>0.01) in geological risk research in the QTP regions included “Sichuan-Tibet Railway”, “Permafrost”, “Landslide”, “Debris flow”, “Climate change”, “Active fracture”, and “Collapse”. The English keywords included “Permafrost”, “Landslide”, “Debris flow”, and “Thermokarst”. All these keywords related to mountain geological hazards and frozen soil distresses. On the other hand, geohazard field surveys indicated that common external geohazards in the QTP regions included collapse (including ice and snow avalanches), landslides, debris flows (including rock debris flows, glacier debris flows, and freeze–thaw mudflows), unstable slopes, and frozen soil distresses. Internal geohazards included high-intensity earthquakes, fault activity, and high ground temperature. Among them, collapses, landslides, debris flows, and unstable slopes had widespread distribution, particularly near active faults and areas prone to high-intensity earthquakes.

In effect, the roadside geological risks covered in this study included collapses, landslides, debris flows, unstable slopes, and frozen soil distresses. Among these hazards, collapses, landslides, debris flows, and unstable slopes were specifically classified as mountain geological hazards.

2.2. Assessment of roadside geological risks

Aforementioned mountain geological hazards are often simplified as point features on small-scale maps. Areas with a dense spatial distribution of these hazards are considered to have higher geological risks. Therefore, this study utilized GIS-based hotspot analysis to identify the spatial clustering of mountain geological hazards. On the other hand, the underlying cause of frozen soil distresses arises from an imbalance among temperature, water, and structure within the frozen soil, leading to reduced stability. Given that temperature plays a significant role in determining frozen soil stability, this study assessed the thermal stability of frozen soil by evaluating the normalized spectral entropy of ground surface temperature on the roadside.

2.2.1. Hotspot analysis of roadside mountain geological hazards

To determine the critical roadside width in relation of hazard occurrences, a statistical analysis was conducted on the number of mountain geological hazards within different widths of the road buffer zone. If both the quantity and severity of hazards need to be considered, equivalent values can be used in place of the number of mountain geological hazards. Herein, an optimized hotspot analysis tool in GIS was used to identify areas with a high frequency of roadside mountain geological hazards (Havas and Resch, 2021; Zhu et al., 2021). This tool aggregates point events into weighted features and automatically corrects for multiple testing and spatial dependence using the False Discovery Rate (FDR) (Wang and Lam, 2020). The tool offers three event data aggregation methods:

(1) Creating grids based on calculated cell sizes and counting the number of points within each grid cell.

(2) Counting the number of points within each provided polygon feature.

(3) Calculating a capture distance using averages and medians nearest neighbors and aggregating the number of points nearby

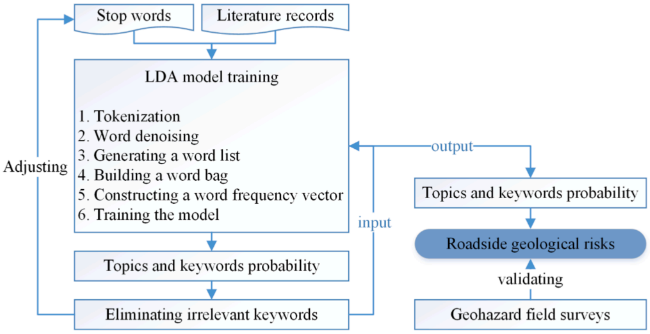

Fig. 3. Process for identifying roadside geological risks in the QTP regions.

Table 2
Topics and keywords probability of roadside geologic risks in the QTP regions.

Topics	Chinese keywords and their probabilities	English keywords and their probabilities
Mountain geological hazards	Landslide (0.0180)	Permafrost (0.0159)
	Debris flow (0.0163)	Landslides (0.0112)
	Climate change (0.0162)	Debris flow (0.0112)
	Collapse (0.0110)	Thermokarst (0.0100)
	Unmanned aerial vehicle (0.0093)	Remote sensing (0.0088)
Frozen soil	Permafrost (0.0212)	Earthquake (0.0073)
	Active fracture (0.0161)	Glaciers (0.0059)
	Developmental characteristics (0.0075)	Precipitation (0.0059)
	Stability (0.0057)	Little ice age (0.0045)
	Seasonal frozen soil (0.0057)	Numerical simulation (0.0045)
Slope failure of linear transportation infrastructure	Sichuan-Tibet Railway (0.0203)	Landslide (0.0137)
	Landscape evolution (0.0077)	InSAR (0.0057)
	Expressway (0.0060)	Debris-covered glaciers (0.0044)
	Highway slope (0.0060)	Climate change (0.0044)
	Landslide (0.0060)	Slope failure (0.0031)

using this distance.

In this study, the third aggregation method was used to generate grid cells and count the number of points within each cell. Subsequently, the Getis-Ord *Gi** statistic was calculated for each grid cell using equations (3)-(5).

$$G_i^* = \frac{\sum_{j=1}^{n}\omega_{ij}x_j - \overline{X}\sum_{j=1}^{n}\omega_{ij}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n}\omega_{ij}^2 - \left(\sum_{j=1}^{n}\omega_{ij}\right)^2\right]}{n-1}}} \tag{3}$$

$$\overline{X} = \frac{\sum_{j=1}^{n}x_j}{n} \tag{4}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n}x_j^2}{n} - (\overline{X})^2} \tag{5}$$

where x_j represents the number of points in grid cell j, ω_{ij} represents the spatial weight between grid cells i and j, and n is the total number of point events.

After calculating the *Gi** statistic, the corresponding *p*-value were obtained. A higher (or lower) *Gi** value indicates a greater clustering degree and thus a higher (or lower) geological risk. To account for multiple testing and spatial dependence, the *p*-values that were statistically significant were sorted in ascending order. The largest *p*-value was then removed from the list based on the FDR

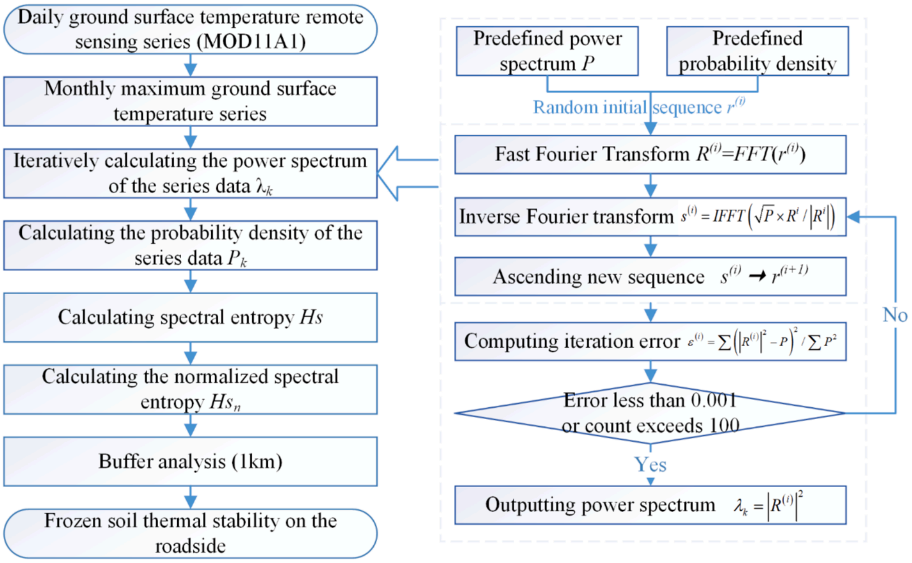

Fig. 4. Assessment process for frozen soil thermal stability on the roadside.

estimation, balancing the effects of multiple tests and spatial relationships by reducing the critical *p*-value threshold.

2.2.2. *Assessment of frozen soil thermal stability on the roadside*

This study utilized the normalized spectral entropy of ground surface temperature to assess the thermal stability of frozen soil in the QTP regions (Zhang et al., 2019). The ground surface temperature data were obtained from the Moderate-Resolution Imaging Spectroradiometer (MODIS) product MOD11A1, downloaded from the United States Geological Survey website (https://lpdaac.usgs.gov/), covering the period from 2010 to 2020. To focus on the roadside areas, a buffer analysis with a width of 1 km was applied, as operationalized in the process shown in Fig. 4.

The formula for calculating spectral entropy based on the monthly maximum ground surface temperature series was given by equation (6).

$$H_S = -\sum_{k=1}^{N} P_k \ln(P_k) \tag{6}$$

where N represents the number of frequencies obtained through Fast Fourier Transform, which is half the length of the time series. P_k denotes the probability of the amplitude value of the kth signal occurring at any position in the signal, i.e., the probability density of the monthly maximum ground surface temperature series. The calculation of P_k was given by the following equation (7).

$$P_k = \frac{|\lambda_k|^2}{\sum_i |\lambda_i|^2} \tag{7}$$

where λ_k represents the power spectrum of the kth signal, derived from the monthly maximum ground surface temperature series for the kth month. The iterative calculation process for obtaining the power spectrum λ_k can be found in Fig. 4.

To obtain the normalized spectral entropy *Hsn*, the spectral entropy was typically normalized to a value between 0 and 1, as shown in the following equation (8). In the time series analysis of ground surface temperature from MOD11A1, a higher value of *Hsn* indicated greater temporal instability of the ground surface temperature and a stronger thermal impact of highways on frozen soil (Sun and Southworth, 2013). The whole calculation process was implemented using the R programming language (Zhang et al., 2019).

$$H_{Sn} = \frac{H_S}{\ln(N)} \tag{8}$$

where $\ln(N)$ is the theoretical maximum of the spectral entropy.

2.3. *Modeling of roadside geological risks*

The objective is to develop mathematical models that quantified the relationship between geological risks (dependent variables) and natural factors and road attributes (independent variables), aiming to predict potential geohazards along roadsides. These models incorporate the influence of road attributes alongside natural factors, departing from traditional mathematical models that considered only natural factors.

For the independent variables, drawing on relevant studies on geohazard susceptibility assessment (Chang et al., 2021; Liang et al., 2021; Loro et al., 2014; Mao et al., 2023), geological risks were found to be associated with factors such as road attributes, natural climate, geological structures, and topography. A detailed indicator system related to these factors is presented in Table 3. For the dependent variables, susceptibility to mountain geological hazards was represented by the probability of such hazards occurring or evolving into hotspots. Similarly, susceptibility to frozen soil distresses was represented by the value of frozen soil thermal stability (i.e., *Hsn*).

The aforementioned influencing factors and risk susceptibilities (i.e., independent and dependent variables) include both

Table 3
A common indicator system for geohazard susceptibility assessment.

Types of geohazards	**Influencing factors**			
	Natural climate	**Geological structures**	**Topography**	**Road attributes**
Mountain geological hazards	■ Precipitation	■ Lithology ■ Fault zones ■ Peak acceleration	■ Elevation ■ Relief degree ■ Slope ■ Aspect ■ Normalized difference vegetation index (NDVI) ■ Water systems	■ Road length ■ Road operation duration ■ Road structure
Frozen soil distresses	■ Precipitation ■ Temperature	/	■ Elevation ■ NDVI ■ Frozen soil types ■ The changes in land cover	■ Road length ■ Road grade ■ Road operation duration ■ Road structure

continuous and categorical variables, and the dependent variables may follow different distributions. Moreover, a single highway may provide multiple data sampling points that exhibit correlations. Building separate relationship models for each data point would be labor-intensive (Zhang et al., 2020; Zhang et al., 2024). Therefore, this study treated individual highways as subject variables and introduced the GEE for geographical data modeling. The relationship between the response variable Y and the predictor variables $\overrightarrow{X}$ can be represented as equations (9)-(11).

$$Y\ \Phi(a, b) \tag{9}$$

$$\mu = \mathrm{E}(Y) \tag{10}$$

$$\mathrm{g}(\mu) = f\left(\overrightarrow{X}\right) \tag{11}$$

where Y was defined as the binary occurrence of mountain geological hazards at a given sampling point or the probability of that point being classified as hotspots for such hazards, as well as the value of *Hsn.* $\Phi(\cdot)$ is a distribution function of the response variable, such as the normal distribution or binomial distribution. Parameters a and b are specific to the chosen distribution function. μ represents the expected value of the response variable. $\mathrm{g}(\cdot)$ is the link function that transforms the marginal expectation of the response variable as a linear combination of covariates, such as the identity function or the logit function. $f\left(\overrightarrow{X}\right)$ is a linear function, where $\overrightarrow{X}$ represents independent variables, such as road attributes, natural climate, geological structures, and topography of the sampling points.

The key to establishing and optimizing GEE model lies in identifying the subject and within-subject variables, and determining the probability distribution, link function, working correlation matrix structure, and independent variables based on the Quasi-likelihood under Independence Model Criterion (QIC) and the Corrected Quasi-likelihood under Independence Model Criterion (QICC) (Kwon et al., 2017). In this study, the subject variable was represented by the road name, while the within-subject variable corresponded to the IDs of the sampling points for each road. The process of establishing and optimizing the GEE using IBM SPSS statistics software involves three key steps:

(1) Selecting an appropriate probability distribution function and link function based on the distribution of the dependent variable.

(2) Specifying the working correlation matrix structure for within-subject cases (i.e., data points on a single highway), considering options such as independent, first-order autoregressive, exchangeable, M−related, or unstructured. Each structure's QIC values were compared to identify the one with the lowest QIC.

(3) Specifying the independent variables and their interaction effects, followed by a comparison of QICC values to select the model with the lowest QICC.

3. Case studies

3.1. Study region

The study region selected for this research was the Western Sichuan Plateau in China (97°-104.5°E, 34°-28°N), situated at the southeastern edge of the QTP (Fig. 5a). The average elevation of the Western Sichuan Plateau exceeds 4,000 m, with higher elevations ranging from 4,500 m to 4,700 m in the northwest and lower elevations ranging from 2,500 m to 3,000 m in the southeast. The climate in this area is characterized as a continental plateau climate, with long hours of sunshine, abundant rainfall (ranging from 289.25 mm to 1111.71 mm annually), and cold temperatures (with an average annual temperature ranging from −14.64℃ to 16.19℃). The water system in the Western Sichuan Plateau is extensive, including major rivers such as the Yangtze River and Yellow River. The Yangtze River system is primarily located in the western, central, and southeastern parts of the region, often forming high mountains and deep valleys with steep cliffs. In contrast, the Yellow River system is mainly found in the central and northeastern parts, characterized by wide floodplains, eroded plains, and residual hills. The main types of bedrock in the Western Sichuan Plateau consist of sedimentary, metamorphic, and igneous rocks, with sedimentary rocks being the most widely distributed, particularly concentrated in the Suomo River area. The geological structure of the Western Sichuan Plateau have been influenced by tectonic movements and can be divided

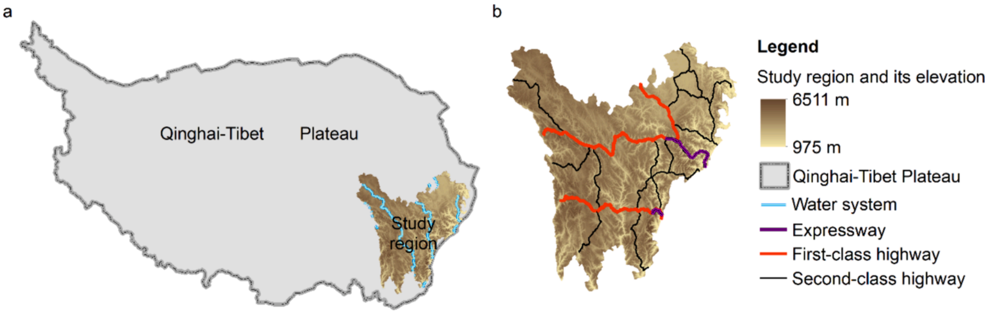

Fig. 5. The geographic location (a) and current road network (b) of the study region.

into three zones from west to east: the Sanjiang fold belt, the Songpan-Ganzi fold belt, and the Longmen Mountain thrust belt, along with the Longmen Mountain foreland basin.

Due to the combined influence of meteorological, hydrological, and geological factors, the Western Sichuan Plateau is characterized by adverse geological conditions (Xu et al., 2020), including collapses, landslides, debris flows, unstable slopes, and seasonal frozen soil. Additionally, the current road network in the Western Sichuan Plateau is limited in quantity and lacks technical standards. As of 2021, the total lengths of expressways, first-class highways, and second-class highways in the region were 511.679 km, 1,293.442 km, and 3,106.370 km, respectively (Fig. 5b).

3.2. Data sources

According to the research methodology of roadside geological risk assessment and modeling, the required data and parameters for the study are shown in Table 4, along with their types and respective sources.

3.3. Hotspot analysis of roadside mountain geological hazards

Based on field surveys and data analysis, the distribution of mountain geological hazards in the study region is illustrated in Fig. 6. The results showed that collapses, landslides, debris flows, and unstable slopes were primarily found in valley areas and along existing highways, with a total of 752, 1,160, 1,578, and 487 instances, respectively.

Then, the statistical analysis was conducted on the number of mountain geological hazards within different road buffer zone widths, namely 25 m, 50 m, 100 m, 250 m, 500 m, 1,000 m, 2,000 m, 3,000 m, and 4,000 m. The analysis revealed that as the road buffer zone width increased, the number of mountain geological hazards tended to increase, albeit at a slower rate. However, when the buffer zone width exceeded 1,000 m, the number of mountain geological hazards on second-class highways reached a stable level. Similarly, when the buffer zone width exceeded 2,000 m, the number of mountain geological hazards on expressways and first-class highways also stabilized. Therefore, the critical buffer zone width was determined to be 2,000 m (Fig. 7).

Next, a GIS-based hotspot analysis was conducted on mountain geological hazards, including collapses, landslides, debris flows, unstable slopes, and overall hazards, within a 2,000 m buffer zone along the roads (Fig. 8). The analysis revealed that mountain geological hazards were more prone to occur at road intersections. Specifically, the hot spots of collapses, landslides, and unstable slopes were consistently concentrated at the junctions of G4217 and G317, as well as G213 and G347. The hot spots of debris flows were concentrated along G350 and its junctions with S210 and G248. Furthermore, the hot spots of overall hazards were concentrated on five highways: G4217, G317, G350, G213, and G347 (at a 99 % confidence interval (CI)), followed by G350, G248, and S211 (at a 90 %-95 % CI). These areas had elevations ranging from 982 m to 4,927 m, NDVI values ranging from 0.416 to 0.896, soil erodibility factor ranging from 0.032 to 0.043 ton•ha•hr(MJ•ha•mm)$^{-1}$, and precipitation erosivity ranging from 768.863 to 1,916.65 MJ•mm (ha•hr)$^{-1}$. On the other hand, the relative cold spots of overall hazards were concentrated on G317 and G227 (at a 90 %-95 % CI).

3.4. Assessment of frozen soil thermal stability on the roadside

The Western Sichuan Plateau is predominantly covered by seasonally frozen ground, which accounts for more than 88.22 % of the regional area. Most of the frozen soil in this area is middle thick (>1m), with sporadic occurrences of thin seasonally frozen ground (<1m) found in the eastern, central, and southwestern regions, along with occasional patches of short-term frozen ground. Some patches of mountain permafrost can be found in the northwest and eastern parts, while predominantly continuous permafrost exists in the northwest (Fig. 9a). The existing roads mainly traverse seasonal and short-term frozen ground.

The results of the frozen soil thermal stability assessment, represented by the *Hsn* index, are shown in Fig. 9b. The average *Hsn* in

Table 4
Types of data and parameters and their sources.

Data/parameters	Types	Sources
Annual precipitation	Raster	The Data Center for Resources and Environmental Science of Chinese Academy of Sciences (https://www.resdc.cn/)
Elevation	Raster	
Water systems	Line features	
NDVI	Raster	
Land cover	Raster	
Mountain geological hazards	Point features	Field surveys, literature review, geological bureau
Fault zones	Line features	
Lithology	Polygon features	
Relief degree	Raster	Elevation
Slope and aspect	Raster	Elevation
Peak acceleration	Polygon features	Seismic ground motion parameters zonation map of China (GB18306-2015)
MOD11A1	Hierarchical Data Format (HDF)	The United States Geological Survey (https://lpdaac.usgs.gov/)
Road grade and road structure	Line features	OpenStreetMap (https://www.openstreetmap.org/)

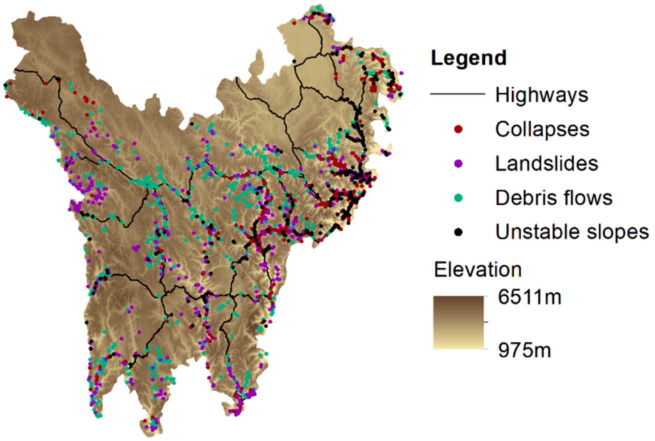

Fig. 6. Distribution of mountain geological hazards in the study region.

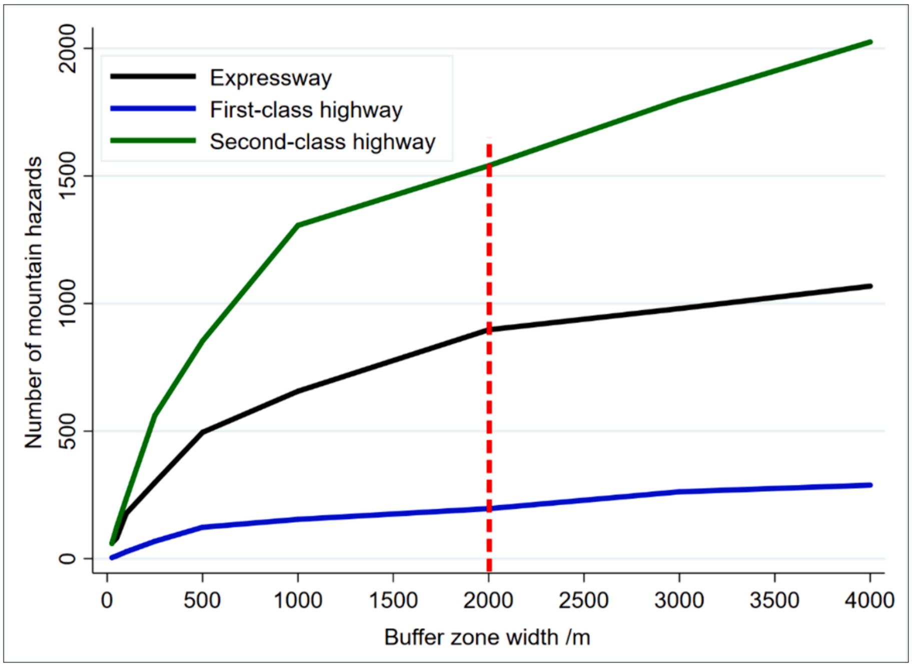

Fig. 7. Number of mountain geological hazards within different buffer zone widths for different road grades.

the study region was found to be 0.469, while the average *Hsn* on the roadside (within a 1 km buffer width) was 0.564. This indicated that the frozen soil thermal stability on the roadside was relatively lower compared to the entire Western Sichuan Plateau region. The northwestern region exhibited higher average *Hsn* values, suggesting a lower frozen soil thermal stability. Conversely, the eastern and southern regions, characterized by lower elevations and higher annual average temperatures, demonstrated higher frozen soil thermal stability.

The average *Hsn* values were further analyzed within the roadside of expressways, first-class highways, and second-class highways, as shown in Fig. 10a. The results indicated that expressways and first-class highways had a similar negative impact on frozen soil thermal stability, which was greater compared to second-class highways. This difference could be attributed to the higher traffic volume and wider embankments of higher-grade highways, which increase heat absorption intensity and extend heat diffusion paths, thereby exerting a greater influence on frozen soil thermal stability.

Furthermore, the average *Hsn* values were analyzed within the roadside of roadbed sections, bridge and culvert sections, and tunnel sections, as shown in Fig. 10b. The results indicated that roadbed sections and bridge and culvert sections had a similar negative impact on frozen soil thermal stability, which was greater compared to tunnel sections. This is primarily due to the higher heat absorption intensity of the surfaces of roadbed sections, resulting in a more pronounced impact on frozen soil thermal stability.

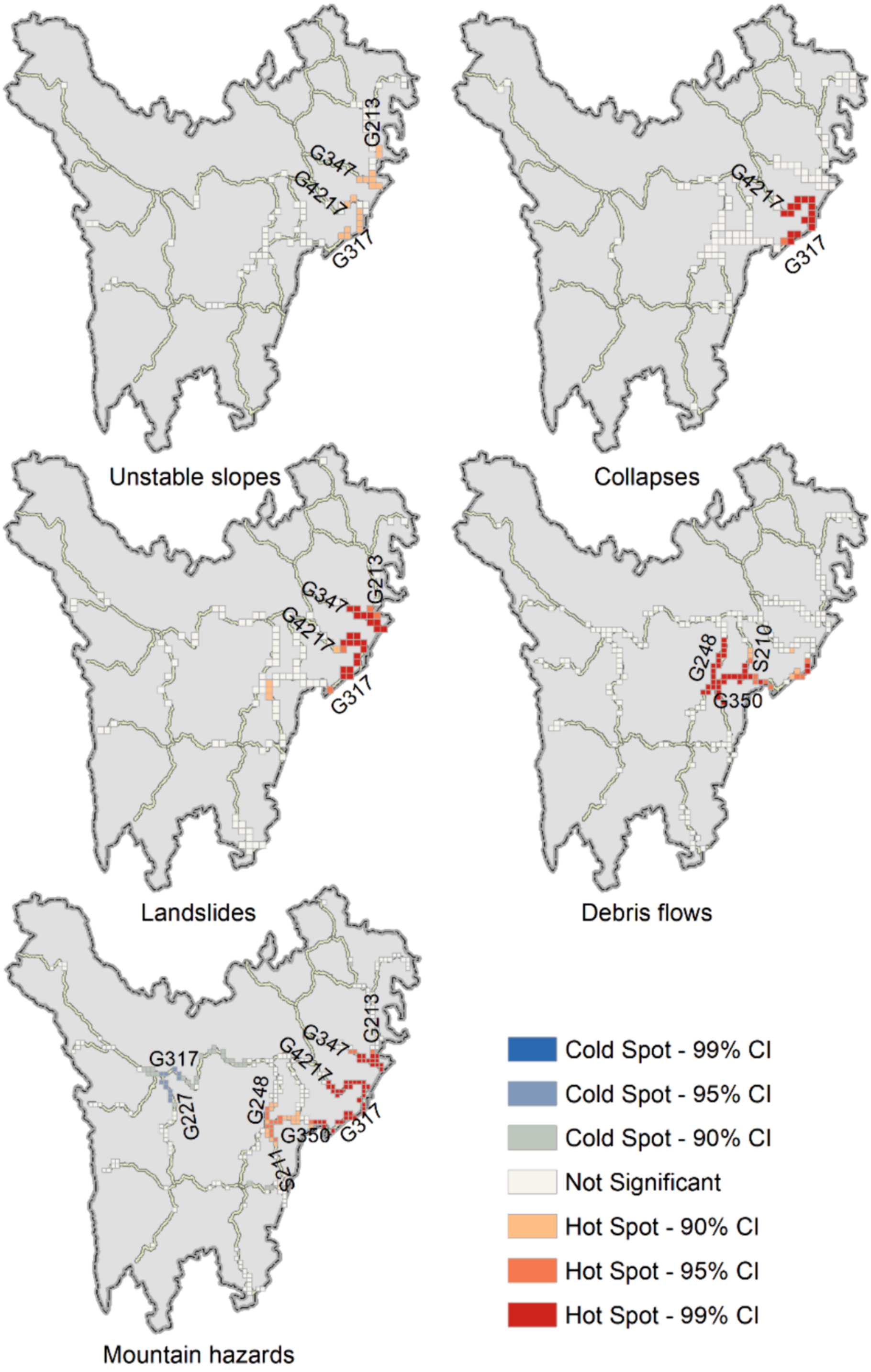

Fig. 8. Hotspot analysis results of roadside mountain geological hazards.

3.5. *GEEs for roadside mountain geological hazards*

In this analysis, several independent variables were considered: road structure (*RS*=1 for roadbed sections; *RS*=2 for tunnel sections; *RS*=3 for bridge and culvert sections), road length (*RL*), road operation duration (*RY*), distance from road centerline (*DC*), elevation (*E*), relief degree (*RD*), slope (*S*), aspect (*A*), *NDVI*, precipitation (*P*), distance from water system (*DW*), lithology (*L*=1 for hard rock formation; *L*=2 for interbedded soft and hard rock; *L*=3 for relatively hard rock formation; *L*=4 for soft rock formation),

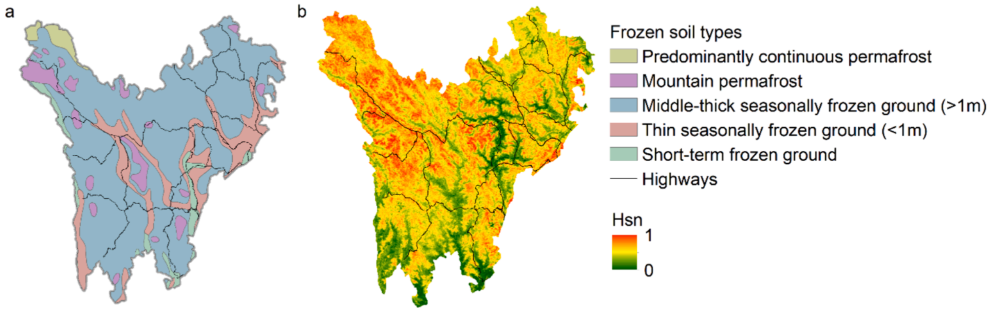

Fig. 9. Distribution of frozen soil (a) and *Hsn* index (b) in the study region.

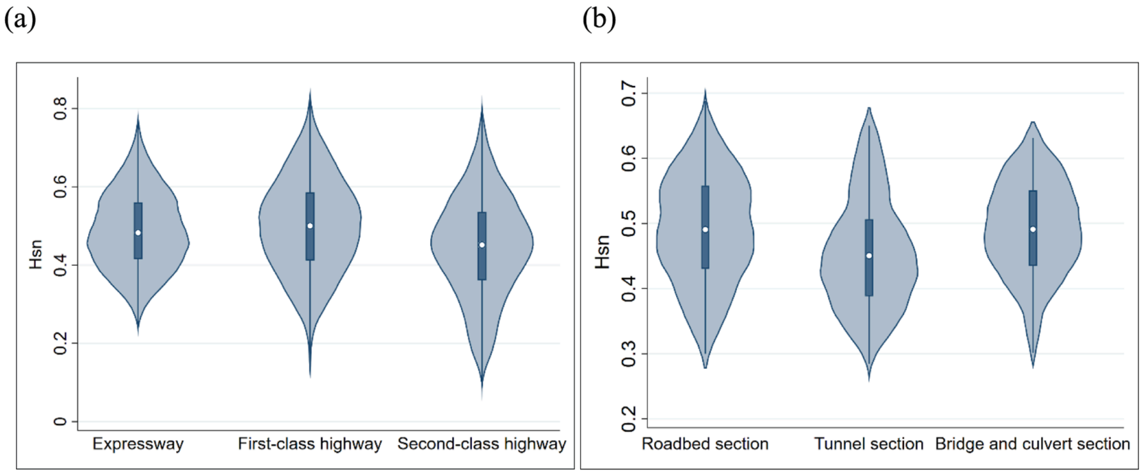

Fig. 10. *Hsn* of different grades of highways (a) and different road structures (b).

Table 5
GEE modeling results for the occurring probability of roadside mountain geological hazards.

Parameters (Independent variables)	Estimated coefficients (*B*)	Standard error	95 % Wald CI for estimated coefficients		Significance of hypothesis tests (*P*)
			Lower	Upper	
Intercept	−30.898	10.411	−51.304	−10.492	0.003
RS_1 = Roadbed sections	17.838	0.747	16.373	19.302	0
RS_2 = Tunnel sections	17.816	0	17.816	17.816	0
RS_3 = Bridge and culvert sections	Control group				
Road length (*RL*)	0	0	0	0	0.805
Road operation duration (*RY*)	0.022	0.005	0.012	0.032	0
Distance from road centerline (*DC*)	−0.001	0	−0.001	−0.001	0
Elevation (*E*)	−0.001	0	−0.002	−0.001	0
Relief degree (*RD*)	0	0	0	0.001	0.023
Slope (*S*)	0.004	0.003	−0.002	0.010	0.177
Aspect (*A*)	−0.001	0	−0.001	0	0
NDVI	−1.647	0.403	−2.436	−0.857	0
Precipitation (*P*)	0.001	0	0	0.001	0.009
Distance from water system (*DW*)	0	0	0	0	0
L_1 = Hard rock formation	−0.455	0.172	−0.792	−0.118	0.008
L_2 = Interbedded soft and hard rock	0.252	0.077	0.102	0.402	0.001
L_3 = Relatively hard rock formation	−0.230	0.086	−0.398	−0.062	0.007
L_4 = Soft rock formation	Control group				
Distance from fault zones (*DF*)	0	0	0	0	0
Peak acceleration (*PA*)	1.589	0.615	0.384	2.793	0.01
Goodness of fit	*QIC*=9090.666; *QICC*=8896.304				

distance from fault zones (*DF*), and peak acceleration (*PA*). The binary occurrence of roadside mountain geological hazards was the dependent variable, where a value of 0 indicated non-occurrence and 1 signified occurrence. The probability distribution, link function, and working correlation matrix structure were set as the binomial distribution, logit function, and independent, respectively. The GEE modeling results are shown in Table 5.

The results indicated that, except for road length (P=0.805) and slope (P=0.177), all other influencing factors were found to be significant at a 95 % confidence level. Among them, the influence of relief degree, distance from water systems, and distance from fault zones was minimal or negligible (B=0). Specifically, compared to bridge and culvert sections, the probability of mountain geological hazards occurring was significantly higher in roadbed sections and tunnel sections (B>0). This may be attributed to the disturbances caused by roadbed sections to the original surface vegetation, soil, and terrain, as well as the increased surface undulation due to steep slopes. Conversely, tunnel sections, although located in geologically prone areas, demonstrated reduced susceptibility to mountain geological hazards due to the adoption of tunnel structures, which mitigated the impact on the route. As road operation duration (B=0.022), peak acceleration (B=1.589), and precipitation (B=0.001) increased, the susceptibility to mountain geological hazards also increased. However, an increase in distance from the road centerline, elevation, slope, and aspect (all with B values of −0.001), as well as an increase in NDVI (B=-1.647), resulted in a decrease in the probability of mountain geological hazards occurring. This indicated that geological disasters were more likely to occur on lower elevation valley slopes. Moreover, compared to soft rock formations, hard rock formations (B=-0.455) and relatively hard rock formations (B=-0.230) reduced the susceptibility to mountain geological hazards, while interbedded soft and hard rock formations (B=0.252) increased the susceptibility to these hazards.

In a similar manner, taking the same independent variables, the Getis-Ord *Gi** statistic from the GIS-based hotspot analysis was the dependent variable (*HH*), where values of 3, 2, 1, and 0 indicated 99 %, 95 %, 90 %, and 0 % CI for hazards becoming hotspots, respectively. The probability distribution, link function, and working correlation matrix structure were set as the multinominal distribution, cumulative logit function, and independent, respectively. The GEE modeling results are shown in Table 6.

The results indicated that road length, elevation, relief degree, NDVI, distance from fault zones, and peak acceleration were significant influencing factors at a 95 % confidence level (P<0.05). The effects of road length, distance from road centerline, distance from water systems, and distance from fault zones were minimal or negligible (B=0). Increased road grade, relief degree, aspect, precipitation, and peak acceleration (B<0) were linked to a higher probability of roadside mountain geological hazards becoming hotspots. This indicated that lower-grade roads in seismic zones with high topographic relief were more likely to become hotspots for mountain geological hazards. However, increased road operation duration, elevation, slope, NDVI, and rock type (B>0) were associated with a decreased probability of these hazards becoming hotspots. This suggested that roadside mountain geological hazards were more likely to occur in areas with lower elevation and sparse vegetation, even if the rock formations were predominantly hard. Protective measures during road operations may help mitigate these hazards. Additionally, compared to bridge and culvert sections, the probability of roadside mountain geological hazards becoming hotspots was significantly higher in roadbed sections (B<0).

The modeling information and results regarding the different roadside mountain geohazards types are detailed in Appendix. The

Table 6
GEE modeling results for the probability of roadside mountain geological hazards classified as hotspots.

Parameters (Independent variables)	**Estimated coefficients (*B*)**	**Standard error**	**95 % Wald CI for estimated coefficients**		**Significance of hypothesis tests (*P*)**
			Lower	**Upper**	
Threshold *HH*=3 vs Control group: *HH*<=2	−17.604	2.974	−23.434	−11.775	0
Threshold *HH*>=2 vs Control group: *HH*<=1	−17.048	2.986	−22.900	−11.194	0
Threshold *HH*>=1 vs Control group: *HH*=0	−16.533	3.016	−22.444	−10.622	0
RS_1 = Roadbed sections	−20.085 [a]				
RS_3 = Bridge and culvert sections	Control group				
Road length (*RL*)	0	0	0	0	0.002
Road grade (*RG*)	−0.204	0.225	−0.645	0.237	0.364
Road operation duration (*RY*)	0.021	0.019	−0.017	0.059	0.275
Distance from road centerline (*DC*)	0	0	0	0	0.962
Elevation (*E*)	0.002	0	0.001	0.002	0
Relief degree (*RD*)	−0.001	0.001	−0.002	0	0.027
Slope (*S*)	0.005	0.005	−0.005	0.014	0.323
Aspect (*A*)	−0.001	0.001	−0.002	0.001	0.310
NDVI	4.406	2.077	0.335	8.477	0.034
Precipitation (*P*)	−0.002	0.002	−0.005	0.001	0.149
Distance from water system (*DW*)	0	0	0	0	0.301
Rock type (*L*)	0.109	0.108	−0.103	0.321	0.313
Distance from fault zones (*DF*)	0	0	0	0	0.001
Peak acceleration (*PA*)	−4.643	2.363	−9.275	−0.011	0.049

Note: [a] represents the parameter estimate at the last iteration. Since Hessian matrix singularity is caused by this parameter, only the estimated coefficient is displayed without other statistics such as standard error, 95% Wald CI, and *P* values. If this parameter is removed, it will have a minimal impact on the modeling results.

findings indicated that while significance levels of the influencing factors varied, the trends associated with these factors were largely consistent across all hazard types. Although some trends may differ, the overall impact on the results was negligible due to the small values of the estimated coefficients.

3.6. GEE for frozen soil thermal stability on the roadside

In this section, several independent variables were analyzed, including road structure (*RS*=1 for roadbed sections; *RS*=2 for tunnel sections; *RS*=3 for bridge and culvert sections), road grade (*RG*=1 for expressways; *RG*=2 for first-class highways; *RG*=3 for second-class highways), road length (*RL*), road operation duration (*RY*), distance from road centerline (*DC*), *NDVI*, land cover change (*LC*=17, 27, 37, 47, 57, 67, 77, representing conversions of farmland, woodland, pastureland, waterbody, built-up land, unexploited land, and road land to road land, respectively), precipitation change (*PC*), annual average temperature change (*TC*), elevation (*E*), and permafrost type (*PT*=1 for predominantly continuous permafrost; *PT*=2 for mountain permafrost; *PT*=3 for middle-thick seasonally frozen ground (>1m); *PT*=4 for thin seasonally frozen ground (<1m); *PT*=5 for short time frozen ground). The dependent variable was the roadside normalized spectral entropy (*Hsn*). The probability distribution, link function, and working correlation matrix structure were set as the normal distribution, identity function, and independent, respectively. The GEE modeling results are shown in Table 7.

The results indicated that several factors, including road grade, road structure, precipitation change, annual average temperature change, and elevation, were found to be significant at a 95 % confidence level. However, the impact of changes in precipitation and elevation was relatively small. Specifically, compared to bridge and culvert sections, roadbed sections exhibited a decrease in frozen soil thermal stability (*B*=0.035), although this effect was not statistically significant (*P*=0.064). In contrast, tunnel sections were found to enhance the frozen soil thermal stability (*B*=-0.038). Additionally, both expressways and first-class highways were found to decrease frozen soil thermal stability compared to second-class highways, although the effect of first-class highways was not statistically significant (*P*=0.154). As temperature, precipitation, and elevation increased, the frozen soil thermal stability decreased (*B*>0). Furthermore, both predominantly continuous permafrost and mountain permafrost were found to have lower thermal stability compared to short time frozen ground and seasonally frozen ground, although this effect was not statistically significant (*P*=0.208).

3.7. Validation of GEEs

To validate the GEEs for roadside mountain geological hazards, receiver operating characteristic curve (ROC) analysis was employed to evaluate the disparity between the mean predicted values (which estimate the probability of roadside mountain geological hazards occurring or classified as hotspots) and the actual values. The ROC serves as a valuable tool for assessing the performance of a classification model by illustrating the relationship between the true positive rate (sensitivity) and the false positive rate (1-specificity). The area under the curve (AUC) of the ROC is often used to quantify the overall performance of the model, with values ranging from 0 to 1. A value closer to 1 indicates better model performance. The ROC analysis indicated that the AUC of the mean predicted values from the GEE, which predicted the probability of roadside mountain geological hazards occurring, was 0.835 (Fig. 11a), demonstrating strong classification performance. Additionally, the AUC scores of the cumulative mean predicted probabilities across the three classifications, derived from the GEE for predicting the probability of roadside mountain geological hazards classified as hotspots, were 0.839, 0.837, and 0.842 (Fig. 11b), respectively, thereby corroborating the model's robust classification capability.

In relation to the GEE for frozen soil thermal stability on the roadside, the efficacy of the model was assessed through residual analysis, given that the dependent variable was continuous. The residual plots revealed that the residuals for the GEE model ranged from −0.4 to 0.4, exhibited normal distributed, and maintained a narrow range of 95 % CI (Fig. 12). The mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE) of the residuals were 0.013, 0.115, 0.092, 23.794 %, and 46.355 %, respectively, indicating that the model had good predictive performance.

4. Conclusions

This study aimed to identify, assess, and model geological risks along roadsides in the QTP regions. The primary contribution is to propose a promising methodology for quantitatively assessing and predicting roadside geological risks, and to explore the synergistic impacts among road attributes and natural factors to geological risk susceptibility. A case study was conducted to validate and examine the proposed approaches, yielding the following primary conclusions.

(1) Common roadside geological risks in the QTP region include mountain geohazards and frozen soil distresses. Specifically, mountain geohazards consist of collapses, landslides, debris flows, and unstable slopes, with an impact distance of around 2,000 m. These hazards are more likely to occur at road intersections, and are primarily concentrated in areas with high soil erodibility and precipitation erosion. Additionally, road construction can significantly diminish the thermal stability of frozen soil, with the effects being more pronounced on higher-grade highways and roadbed sections due to their greater heat absorption.

(2) In addition to traditional natural factors such as slope, vegetation, rock hardness, and seismic activity, lower-grade highways, roadbed sections, and closer distance from the road centerline are also linked to a higher probability of roadside mountain geohazards. For landslides and unstable slopes, the occurring probability decreases with longer road operation duration. However, the opposite is true for collapses and debris flows, where the risk increases over time. These findings may be related to the simpler slope protection measures on lower-grade highways, as designers typically prioritize the management of landslides and unstable slopes. In contrast,

Table 7
GEE modeling results for frozen soil thermal stability on the roadside.

Parameters (Independent variables)	Estimated coefficients (*B*)	Standard error	95 % Wald CI for estimated coefficients		Significance of hypothesis tests (*P*)
			Lower	Upper	
Intercept	0.335	0.057	0.225	0.446	0
RS_1 = Roadbed sections	0.035	0.019	−0.002	0.071	0.064
RS_2 = Tunnel sections	−0.038	0.002	−0.042	−0.035	0
RS_3 = Bridge and culvert sections	Control group				
RG_1 = Expressways	0.070	0.031	0.010	0.130	0.022
RG_2 = First-class highways	0.032	0.023	−0.012	0.077	0.154
RG_3 = Second-class highways	Control group				
Road length (*RL*)	0	0	0	0	0.721
Road operation duration (*RY*)	0.002	0.002	−0.003	0.006	0.404
Distance from road centerline (*DC*)	0	0	0	0	0.166
NDVI	−0.017	0.064	−0.141	0.108	0.795
Land cover change (*LC*)	0	0	0	0	0.756
Precipitation change (*PC*)	0	0	0	0	0.005
Annual average temperature change (*TC*)	0.059	0.013	0.034	0.085	0
Elevation (*E*)	0	0	0	0	0
Permafrost type (*PT*)	−0.010	0.008	−0.026	0.006	0.208
Goodness of fit	*QIC*=985.750; *QICC*=140.722				

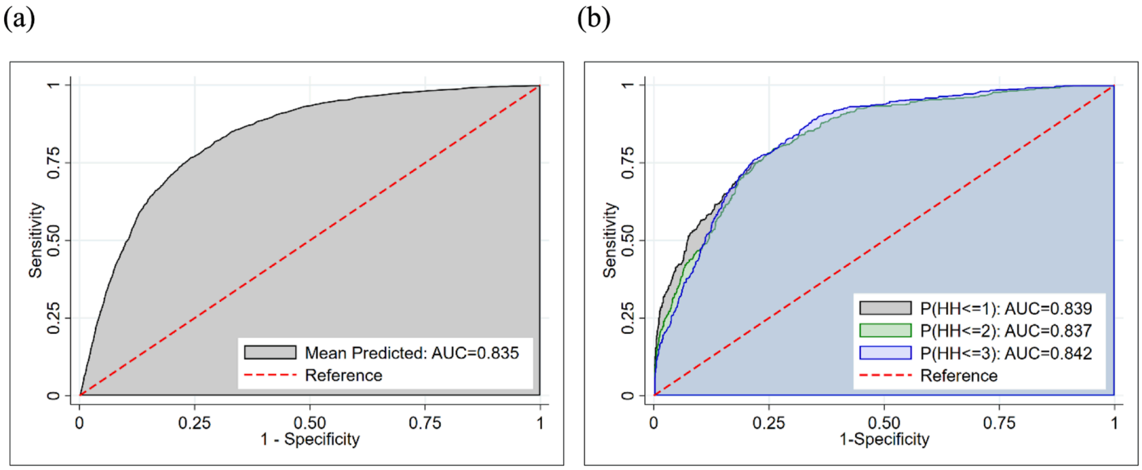

Fig. 11. ROC analysis of GEEs for predicting the probability of roadside mountain geological hazards: (a) occurring and (b) classification as hotspots.

debris flows are more closely associated with short-duration heavy rainfall, rendering them more unpredictable in nature.

(3) Higher-grade highways, roadbed sections, elevated temperatures, increased precipitation, and higher elevations significantly reduce the thermal stability of frozen soil. Conversely, areas with higher thermal stability are mainly found in valley regions, making them suitable for road alignment.

Thus, the policy implications derived from this study are as follows. First, during the road design phase, it is essential to evaluate road schemes based on longitudinal slope and earthwork volume to minimize alterations to surface slope and vegetation, as excessive excavation and embankment can increase mountain geohazard risks. To prevent frozen soil distresses such as frost heave in road embankments, measures such as soil improvement, moisture control, passive insulation, and structural support should be implemented. Second, during the road construction phase, emphasis should be placed on adopting green construction techniques, such as reducing the construction duration and using topsoil for backfill. Effective measures should also be taken to restore vegetation in soil excavation and disposal areas. Third, during the road operational phase, it is important to focus on disaster prevention measures for roadbed sections and to safeguard tunnel entrances and exits against mountain geohazards, implementing emergency response plans as necessary.

Overall, this study focuses on the susceptibility to geohazard risks on roadside areas. Unlike previous studies that only considered natural attributes, this study also takes into account the influence of route characteristics, including route length, structure, and operation duration. The findings contribute to the planning of highway and railway corridors and the mitigation of roadside geological risks. However, the study has some limitations. First, due to the lower resolution of the raster map, the road attributes analyzed herein

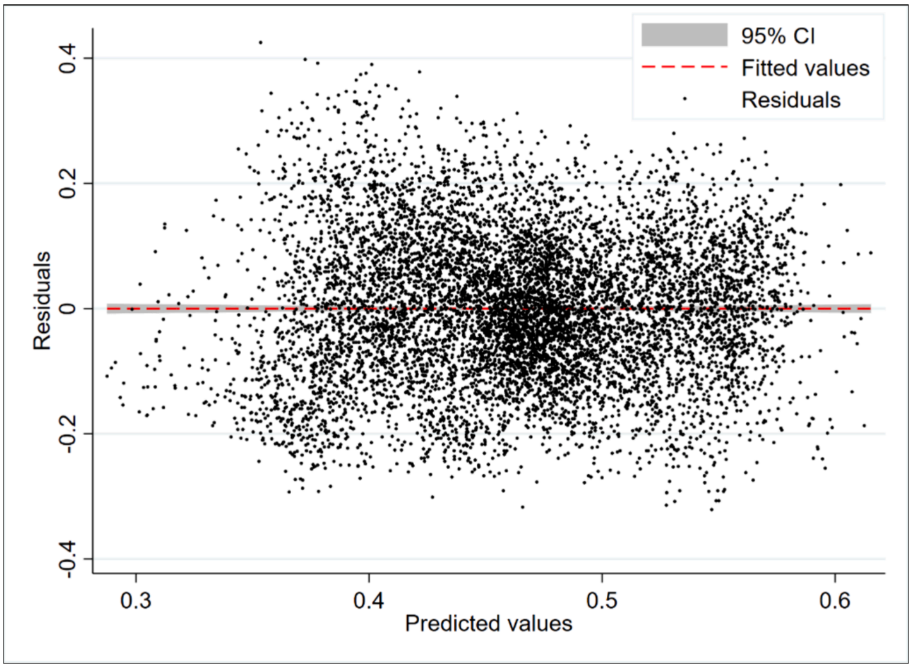

Fig. 12. Residual plots from the GEE for predicting frozen soil thermal stability on the roadside.

are limited to road grade, structure, length, and operation duration. Future research should utilize geospatial data with higher accuracy (≤30 m) to further investigate the coupling relationship between engineering indicators, such as slope steepness, curve radius, and longitudinal slope, and the geological environment. Second, this study may not have comprehensively considered the geological risks in the QTP region, potentially overlooking risks such as high ground temperatures and soft soil risks.

CRediT authorship contribution statement

Hong Zhang: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Xin Xu:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis. **Chi Zhang:** Supervision, Resources, Investigation. **Hong-Zhi Yang:** Validation, Resources. **Yiik Diew Wong:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by National Natural Science Foundation of China (42407646), Inner Mongolia Natural Science Foundation Project, China (2024QN05034 & 2024QN05054), and Inner Mongolia University Young Academic Talent Research Initiation Project, China (10000-23112101/068). The authors thank the Data Center for Resources and Environmental Science of Chinese Academy of Sciences for providing raster maps of the land use, elevation, precipitation, etc.

Appendix

Table A1
GEEs modeling information and results for the occurring probability of different roadside mountain geological hazards.

Types of roadside mountain geological hazards	**Parameters**	***B***	**Standard error**	**95 % Wald CI**		***P***
				Lower	**Upper**	
Collapses	Intercept	−22.026	1.601	−25.163	−18.889	0
	RS_1 = Roadbed sections	20.088 [a]				
	RS_2 = Tunnel sections	−1.578	1.270	−4.066	0.910	0.214
	RS_3 = Bridge and culvert sections	Control group				
	Road length (*RL*)	0	0	0	0	0.414
	Road grade (*RG*)	0.454	0.167	0.245	0.663	0
	Road operation duration (*RY*)	0.003	0.009	−0.015	0.022	0.748
	Distance from road centerline (*DC*)	0	0	−0.001	0	0
	Elevation (*E*)	−0.002	0	−0.002	−0.001	0
	Relief degree (*RD*)	0	0	−0.001	0.001	0.751
	Slope (*S*)	−0.004	0.006	−0.017	0.008	0.517
	Aspect (*A*)	0.001	0.001	−0.001	0.002	0.401
	NDVI	0.975	0.977	−0.940	2.890	0.318
	Precipitation (*P*)	0.002	0.001	0.001	0.003	0.008
	Distance from water system (*DW*)	0	0	0	0	0.515
	L_1 = Hard rock formation	1.784	0.644	0.521	3.047	0.006
	L_2 = Interbedded soft and hard rock	1.590	0.574	0.466	2.715	0.006
	L_3 = Relatively hard rock formation	1.697	0.652	0.419	2.975	0.009
	L_4 = Soft rock formation	Control group				
	Distance from fault zones (*DF*)	0	0	0	0	0.761
	Peak acceleration (*PA*)	−2.479	1.372	−5.169	0.210	0.071
	Goodness of fit	*QIC*=1292.879; *QICC*=1303.566				
Landslides	Intercept	3.255	1.059	1.180	5.331	0.002
	RS_1 = Roadbed sections	−0.510	0.375	−1.244	0.224	0.173
	RS_2 = Tunnel sections	−27.169 [a]				
	RS_3 = Bridge and culvert sections	Control group				
	Road length (*RL*)	0	0	0	0	0.070
	Road grade (*RG*)	−0.129	0.066	−0.259	0.001	0.052
	Road operation duration (*RY*)	−0.022	0.010	−0.041	−0.002	0.027
	Distance from road centerline (*DC*)	0	0	0	0	0.293
	Elevation (*E*)	−0.001	0	−0.001	−0.001	0
	Relief degree (*RD*)	0.001	0.001	0	0.002	0.134
	Slope (*S*)	0.003	0.005	−0.006	0.012	0.562
	Aspect (*A*)	−0.001	0	−0.001	0	0.156
	NDVI	−1.191	0.773	−2.705	0.324	0.123
	Precipitation (*P*)	0	0.001	−0.002	0.001	0.458
	Distance from water system (*DW*)	0	0	0	0	0.453
	L_1 = Hard rock formation	0.347	0.494	−0.620	1.315	0.482
	L_2 = Interbedded soft and hard rock	0.539	0.510	−0.461	1.539	0.291
	L_3 = Relatively hard rock formation	0.274	0.518	−0.741	1.288	0.597
	L_4 = Soft rock formation	Control group				
	Distance from fault zones (*DF*)	0	0	0	0	0.181
	Peak acceleration (*PA*)	−0.575	1.272	−3.068	1.918	0.651
	Goodness of fit	*QIC*=2305.777; *QICC*=2309.841				
Debris flows	Intercept	−19.511	0.739	−20.960	−18.063	0
	RS_1 = Roadbed sections	20.684 [a]				
	RS_2 = Tunnel sections	−1.193	0.971	−3.096	0.710	0.219
	RS_3 = Bridge and culvert sections	Control group				
	Road length (*RL*)	0	0	0	0	0.272
	Road grade (*RG*)	0.154	0.076	0.005	0.304	0.043
	Road operation duration (*RY*)	0.012	0.008	−0.004	0.028	0.147
	Distance from road centerline (*DC*)	0	0	0	0	0.094
	Elevation (*E*)	−0.002	0	−0.002	−0.001	0
	Relief degree (*RD*)	0.001	0	0	0.001	0.178
	Slope (*S*)	−0.004	0.007	−0.017	0.009	0.551
	Aspect (*A*)	0	0	−0.001	0	0.263
	NDVI	1.369	0.655	0.085	2.653	0.037
	Precipitation (*P*)	0.001	0.001	0	0.002	0.241
	Distance from water system (*DW*)	0	0	0	0	0.379
	L_1 = Hard rock formation	−0.245	0.194	−0.625	0.135	0.207
	L_2 = Interbedded soft and hard rock	−0.004	0.213	−0.421	0.413	0.985
	L_3 = Relatively hard rock formation	−0.084	0.296	−0.663	0.496	0.777

(*continued on next page*)

Table A1 (*continued*)

Types of roadside mountain geological hazards	Parameters	*B*	Standard error	95 % Wald CI		*P*
				Lower	Upper	
	L_4 = Soft rock formation	Control group				
	Distance from fault zones (*DF*)	0	0	0	0	0.367
	Peak acceleration (*PA*)	−0.861	1.160	−3.134	1.413	0.458
	Goodness of fit	*QIC*=4247.450; *QICC*=4230.857				
Unstable slopes	Intercept	−20.721	1.299	−23.267	−18.175	0
	RS_1 = Roadbed sections	23.021 [a]				
	RS_2 = Tunnel sections	−0.120	0.893	−1.869	1.629	0.893
	RS_3 = Bridge and culvert sections	Control group				
	Road length (*RL*)	0	0	0	0	0.417
	Road grade (*RG*)	0.073	0.107	−0.137	0.283	0.466
	Road operation duration (*RY*)	−0.001	0.012	−0.024	0.022	0.926
	Distance from road centerline (*DC*)	0	0	−0.001	0	0.001
	Elevation (*E*)	−0.001	0	−0.002	−0.001	0
	Relief degree (*RD*)	0	0	−0.001	0.001	0.483
	Slope (*S*)	0.002	0.005	−0.008	0.013	0.689
	Aspect (*A*)	−0.001	0	−0.001	0	0.146
	NDVI	−0.626	0.999	−2.584	1.331	0.531
	Precipitation (*P*)	0.001	0.001	−0.001	0.003	0.396
	Distance from water system (*DW*)	0	0	0	0	0.965
	L_1 = Hard rock formation	0.406	0.275	−0.134	0.946	0.140
	L_2 = Interbedded soft and hard rock	0.424	0.336	−0.233	1.082	0.206
	L_3 = Relatively hard rock formation	0.610	0.339	−0.055	1.274	0.072
	L_4 = Soft rock formation	Control group				
	Distance from fault zones (*DF*)	0	0	0	0	0.182
	Peak acceleration (*PA*)	−4.330	2.201	−8.643	−0.017	0.049
	Goodness of fit	*QIC*=1528.825; *QICC*=1533.926				
Model information						
Dependent variable		The binary occurrence of hazards at sampling points				
Probability distribution		Binomial				
Link function		Logit				
Working correlation matrix structure		Independent				

Table A2

GEEs modeling information and results for the probability of different roadside mountain geological hazards classified as hotspots.

Types of roadside mountain geological hazards	Parameters	*B*	Standard error	95 % Wald CI		*P*
				Lower	Upper	
Collapses	Threshold	−6.128	3.202	−12.403	0.147	0.056
	HH=3 vs Control group: *HH*<=2					
	Threshold	−5.477	3.183	−11.716	0.763	0.085
	HH>=2 vs Control group: *HH*<=1					
	Threshold	−4.860	3.148	−11.031	1.310	0.123
	HH>=1 vs Control group: *HH*=0					
	Road length (*RL*)	0	0	0	0	0.217
	Road grade (*RG*)	−0.860	0.228	−1.307	−0.413	0
	Road operation duration (*RY*)	0.049	0.028	−0.006	0.103	0.080
	Distance from road centerline (*DC*)	0	0	0	0.001	0.450
	Elevation (*E*)	0.001	0	0	0.002	0.033
	Relief degree (*RD*)	−0.002	0.001	−0.004	0	0.025
	Slope (*S*)	0.014	0.008	−0.002	0.030	0.085
	Aspect (*A*)	−0.002	−0.004	0	5.485	0.019
	NDVI	3.268	2.116	−0.880	7.415	0.123
	Precipitation (*P*)	−0.006	0.003	−0.011	−0.002	0.010
	Distance from water system (*DW*)	0	0	0	0	0.348
	Rock type (*L*)	0.269	0.246	−0.214	0.751	0.275
	Distance from fault zones (*DF*)	0	0	0	0	0
	Peak acceleration (*PA*)	−9.811	4.506	−18.642	−0.980	0.029
Landslides	Threshold	−15.080	3.214	−21.380	−8.780	0
	HH=3 vs Control group: *HH*<=2					
	Threshold	−14.746	3.198	−21.013	−8.479	0
	HH>=2 vs Control group: *HH*<=1					
	Threshold	−14.246	3.236	−20.589	−7.903	0
	HH>=1 vs Control group: *HH*=0					
	RS_1 = Roadbed sections	−21.397 [a]				
	RS_3 = Bridge and culvert sections	Control group				
	Road length (*RL*)	0	0	0	0	0.001

(*continued on next page*)

Table A2 (*continued*)

Types of roadside mountain geological hazards	Parameters	*B*	Standard error	95 % Wald CI		*P*
				Lower	Upper	
	Road grade (*RG*)	0.111	0.225	−0.330	0.553	0.621
	Road operation duration (*RY*)	−0.007	0.031	−0.068	0.055	0.835
	Distance from road centerline (*DC*)	0	0	0	0.001	0.748
	Elevation (*E*)	0.002	0.001	0.001	0.003	0
	Relief degree (*RD*)	−0.001	0.001	−0.002	0	0.125
	Slope (*S*)	0.015	0.011	−0.007	0.038	0.185
	Aspect (*A*)	0	0.002	−0.004	0.003	0.901
	NDVI	2.201	2.867	−3.418	7.820	0.443
	Precipitation (*P*)	0.001	0.001	−0.001	0.003	0.437
	Distance from water system (*DW*)	0	0	0	0	0.612
	Rock type (*L*)	0.266	0.087	0.095	0.436	0.002
	Distance from fault zones (*DF*)	0	0	0	0	0.068
	Peak acceleration (*PA*)	−8.971	1.445	−11.803	−6.139	0
Debris flows	Threshold *HH*=3 vs Control group: *HH*<=2	1.534	2.982	−4.311	7.379	0.607
	Threshold *HH*>=2 vs Control group: *HH*<=1	2.348	2.991	−3.514	8.211	0.432
	Threshold *HH*>=1 vs Control group: *HH*=0	2.913	3.057	−3.079	8.904	0.341
	Road length (*RL*)	0	0	0	0	0.205
	Road grade (*RG*)	−0.221	0.246	−0.704	0.262	0.370
	Road operation duration (*RY*)	0.012	0.018	−0.022	0.046	0.485
	Distance from road centerline (*DC*)	0	0	−0.001	0	0.218
	Elevation (*E*)	0.002	0	0.001	0.002	0
	Relief degree (*RD*)	−0.001	0.001	−0.003	0	0.101
	Slope (*S*)	−0.003	0.009	−0.022	0.015	0.719
	Aspect (*A*)	0	0.001	−0.002	0.002	0.876
	NDVI	6.102	1.828	2.518	9.685	0.001
	Precipitation (*P*)	−0.004	0.001	−0.006	−0.001	0.010
	Distance from water system (*DW*)	0	0	0	0	0.266
	Rock type (*L*)	0.077	0.095	−0.109	0.264	0.416
	Distance from fault zones (*DF*)	0	0	0	0	0
	Peak acceleration (*PA*)	−1.128	3.877	−8.726	6.469	0.771
Unstable slopes	Threshold *HH*=3 vs Control group: *HH*<=2	7.070	4.224	−1.209	15.348	0.094
	Threshold *HH*>=2 vs Control group: *HH*<=1	7.767	4.225	−0.513	16.047	0.066
	Threshold *HH*>=1 vs Control group: *HH*=0	8.416	4.216	0.152	16.679	0.046
	Road length (*RL*)	0	0	0	0	0.493
	Road grade (*RG*)	−0.221	0.181	−0.575	0.133	0.221
	Road operation duration (*RY*)	0.061	−0.030	0.002	0.119	0.042
	Distance from road centerline (*DC*)	0	0	−0.001	0.001	0.931
	Elevation (*E*)	0.003	0.001	0.002	0.004	0
	Relief degree (*RD*)	0	0.001	−0.002	0.001	0.700
	Slope (*S*)	0.038	0.011	0.017	0.060	0
	Aspect (*A*)	−0.002	0.001	−0.005	0	0.094
	NDVI	5.363	3.202	−0.912	11.639	0.094
	Precipitation (*P*)	−0.001	0.004	−0.008	0.006	0.809
	Distance from water system (*DW*)	0	0	0	0	0.263
	Rock type (*L*)	−0.097	0.178	−0.447	0.252	0.585
	Distance from fault zones (*DF*)	0	0	0	0	0.091
	Peak acceleration (*PA*)	−5.246	3.858	−12.808	2.316	0.174
Model information						
Dependent variable		The probability of hazards classified as hotspots at sampling points				
Probability distribution		Multinominal				
Link function		Cumulative logit				
Working correlation matrix structure		Independent				

References

Abdulla, M.B., Costa, A.L., Sousa, R.L., 2018. Probabilistic identification of subsurface gypsum geohazards using artificial neural networks. Neural Comput. & Applic. 29, 1377–1391. https://doi.org/10.1007/s00521-016-2655-3.

Amatya, P., Kirschbaum, D., Stanley, T., 2019. Use of very high-resolution optical data for landslide mapping and susceptibility analysis along the Karnali Highway, Nepal. Remote Sens. (Basel) 11, 23. https://doi.org/10.3390/rs11192284.

Ambrosi, C., Strozzi, T., Scapozza, C., Wegmüller, U., 2018. Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on Earth-Observation data. Eng. Geol. 237 https://doi.org/10.1016/j.enggeo.2018.02.020.

Banks, G.C., Woznyj, H.M., Wesslen, R.S., Ross, R.L., 2018. A review of best practice recommendations for text analysis in R (and a User-Friendly App). J. Bus. Psychol. 33, 445–459. https://doi.org/10.1007/s10869-017-9528-3.

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A.M., Nunes, C., Jatowt, A., 2018. A text feature based automatic keyword extraction method for single documents, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-319-76941-7_63.

Chang, M., Cui, P., Dou, X.Y., Su, F.H., 2021. Quantitative risk assessment of landslides over the China-Pakistan economic corridor. Int. J. Disaster Risk Reduct. 63 https://doi.org/10.1016/j.ijdrr.2021.102441.

Cui, P., Ge, Y.G., Li, S.J., Li, Z.H., Xu, X.W., Zhou, G.G.D., Chen, H.Y., Wang, H., Lei, Y., Zhou, L.B., Yi, S.J., Wu, C.H., Guo, J., Wang, Q., Lan, H.X., Ding, M.T., Ren, J. J., Zeng, L., Jiang, Y.J., Wang, Y., 2022. Scientific challenges in disaster risk reduction for the Sichuan-Tibet Railway. Eng. Geol. 309 https://doi.org/10.1016/j.enggeo.2022.106837.

Dai, L.Y., Zhu, M.C., Ma, Z.G., He, Z.Y., Jiang, H., Li, P.S., Zhang, X.B., Shi, J.B., Chen, K., Weng, T., Zheng, Z.Z., Liu, Q., Ieee, 2020. Recommendation of similar landslide cases based on landslide profile, in: 17th IEEE International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). Univ Elect Sci & Technol China, Chengdu, PEOPLES R CHINA, pp. 50–53. 10.1109/iccwamtip51612.2020.9317346.

Du, Q.S., Li, G.Y., Chen, D., Zhou, Y., Qi, S.S., Wu, G., Chai, M.T., Tang, L.Y., Jia, H.L., Peng, W.L., 2021. SBAS-InSAR-based analysis of surface deformation in the Eastern Tianshan Mountains, China. Front. Earth Sci. (Lausanne) 9, 16. https://doi.org/10.3389/feart.2021.729454.

Fortier, R., LeBlanc, A.M., Yu, W.B., 2011. Impacts of permafrost degradation on a road embankment at Umiujaq in Nunavik (Quebec), Canada. Can. Geotech. J. 48, 720–740. https://doi.org/10.1139/t10-101.

Gao, Z.M., Ding, M.T., Huang, T., Liu, X.W., Hao, Z., Hu, X.W., Xi, C.J., 2022. Landslide risk assessment of high-mountain settlements using Gaussian process classification combined with improved weight-based generalized objective function. Int. J. Disaster Risk Reduct. 67 https://doi.org/10.1016/j.ijdrr.2021.102662.

Ghias, M.S., Therrien, R., Molson, J., Lemieux, J.M., 2017. Controls on permafrost thaw in a coupled groundwater-flow and heat-transport system: Iqaluit Airport, Nunavut, Canada. Hydrgeol. J. 25, 657–673. https://doi.org/10.1007/s10040-016-1515-7.

Havas, C., Resch, B., 2021. Portability of semantic and spatial-temporal machine learning methods to analyse social media for near-real-time disaster monitoring. Nat. Hazards. https://doi.org/10.1007/s11069-021-04808-4.

He, F., Liu, C.X., Liu, H.J., 2021. Integration and fusion of geologic hazard data under deep learning and big data analysis technology. Complexity 2021, 10. https://doi.org/10.1155/2021/2871770.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L., 2019. Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. Multimed. Tools Appl. 78, 15169–15211. https://doi.org/10.1007/s11042-018-6894-4.

Jia, Y., Liu, J., Guo, L., Deng, Z., Li, J., Zheng, H., 2021. Locomotion of slope geohazards responding to climate change in the qinghai-tibetan plateau and its adjacent regions. Sustainability (Switzerland) 13. https://doi.org/10.3390/su131910488.

Jiao, Z., Xu, Z., Guo, R., Zhou, Z., Jiang, L., 2023. Potential of multi-temporal InSAR for detecting retrogressive thaw slumps: A case of the Beiluhe region of the Tibetan Plateau. Int. J. Disaster Risk Sci. 14 https://doi.org/10.1007/s13753-023-00505-x.

Justicia De La Torre, C., Sánchez, D., Blanco, I., Martín-Bautista, M.J., 2018. Text mining: Techniques, applications, and challenges. Int. J. Uncertainty Fuzziness Knowldege-Based Syst. 26 https://doi.org/10.1142/S0218488518500265.

Kusumawati, N.A., Putra, I.D., Wijaya, I., Indrawan, I.G.B., Bandung Inst Technol, F.E.S., Technol, 2017. Integrated remote sensing and geological observation in identifying landslide triggering factors: A case study of landslide hazard of Hargotirto Subdistrict, Kulon Progo, DI Yogyakarta, in: 7th Annual Symposium on Earthquake and Related Geohazard Research for Disaster Risk Reduction. Bandung Inst Technol, Fac Earth Sci & Technol, Bandung, INDONESIA. 10.1063/1.5047349.

Kwon, Y., Choi, Y.G., Park, T., Ziegler, A., Paik, M.C., 2017. Generalized estimating equations with stabilized working correlation structure. Comput. Stat. Data Anal. 106, 1–11. https://doi.org/10.1016/j.csda.2016.08.016.

Li, G., Tang, M.G., Zhang, M.L., Peng, D.L., Zhao, H.L., Zhou, J., 2023. Slope stability under the influence of irrigation and frozen stagnant water effect in Heifangtai. Bull. Eng. Geol. Environ. 82, 18. https://doi.org/10.1007/s10064-023-03264-7.

Li, A., Xia, C., Bao, C., Yin, G., 2019. Using MODIS land surface temperatures for permafrost thermal modeling in beiluhe basin on the Qinghai-Tibet plateau. Sensors (Switzerland) 19. https://doi.org/10.3390/s19194200.

Liang, S.Z.X., Chen, D., Li, D.H., Qi, Y.C., Zhao, Z.F., 2021. Spatial and temporal distribution of geologic hazards in Shaanxi Province. Remote Sens. (Basel) 13, 16. https://doi.org/10.3390/rs13214259.

Lin, D.M., Chen, P., Ma, J.L., Zhao, Y., Xie, T., Yuan, R.M., Li, L., 2019. Assessment of slope construction risk uncertainty based on index importance ranking. Bull. Eng. Geol. Environ. 78, 4217–4228. https://doi.org/10.1007/s10064-018-1387-2.

Liu, W.B., Yu, W.B., Hu, D., Lu, Y., Chen, L., Yi, X., Han, F.L., 2019. Crack damage investigation of paved highway embankment in the Tibetan Plateau permafrost environments. Cold Reg. Sci. Technol. 163, 78–86. https://doi.org/10.1016/j.coldregions.2019.05.003.

Liu, Q., Zhang, B., Tang, A., 2023. Landslide risk of regional roads: Consider the road mileage of expected losses. Transp. Res. D Transp. Environ. 120 https://doi.org/10.1016/j.trd.2023.103771.

Loro, M., Arce, R.M., Ortega, E., Martin, B., 2014. Road-corridor planning in the EIA procedure in Spain. A review of case studies. Environ. Impact Assess. Rev. 44, 11–21. https://doi.org/10.1016/j.eiar.2013.08.005.

Lu, N., Wayllace, A., Oh, S., 2013. Infiltration-induced seasonally reactivated instability of a highway embankment near the Eisenhower Tunnel, Colorado, USA. Eng. Geol. 162, 22–32. https://doi.org/10.1016/j.enggeo.2013.05.002.

Luo, S.R., Feng, G.C., Xiong, Z.Q., Wang, H.Y., Zhao, Y.G., Li, K.F., Deng, K.L., Wang, Y.X., 2021. An improved method for automatic identification and assessment of potential geohazards based on MT-InSAR measurements. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13173490.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Haeussler, T., Schmid-Petri, H., Adam, S., 2018. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Commun. Methods Meas. 12, 93–118. https://doi.org/10.1080/19312458.2018.1430754.

Makarieva, O., Nesterova, N., Lebedeva, L., Sushansky, S., 2018. Water balance and hydrology research in a mountainous permafrost watershed in upland streams of the Kolyma River, Russia: a database from the Kolyma Water-Balance Station, 1948–1997. Earth Syst. Sci. Data 10, 689–710. https://doi.org/10.5194/essd-10-689-2018.

Mao, G., Duan, X., Niu, Z., Xu, J., Xiao, X., Huang, X., Chen, H., Mehr, F., Moti, R., Qiao, Z., 2023. Application of source-sink theory and MCR model to assess hydrochemical change risk in Lhasa River basin, Tibet, China. Environ. Impact Assess. Rev. 101 https://doi.org/10.1016/j.eiar.2023.107124.

Mihalcea, R., Tarau, P., 2004. TextRank: Bringing order into texts, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004 - A Meeting of SIGDAT, a Special Interest Group of the ACL Held in Conjunction with ACL 2004.

Ni, J., Wu, T., Zhu, X., Chen, J., Wu, X., Hu, G., Zou, D., Li, R., Du, Y., 2022. Quantifying the relationship between human activities intensity and thawing hazards of the frozen ground on the Qinghai-Tibet Plateau. Front. Earth Sci. (Lausanne) 10. https://doi.org/10.3389/feart.2022.845873.

Niu, H.T., 2020. Smart safety early warning model of landslide geological hazard based on BP neural network. Saf. Sci. 123 https://doi.org/10.1016/j.ssci.2019.104572.

Onan, A., 2018. Biomedical text categorization based on ensemble pruning and optimized topic modelling. Comput. Math. Methods Med. 2018 https://doi.org/10.1155/2018/2497471.

Onder, H., 2016. Comparative study of generalized estimating equations and logistic regressions on different sample sizes and correlation levels. Commun. Stat.-Simulation Comput. 45, 3528–3533. https://doi.org/10.1080/03610918.2015.1010000.

Ourang, C.J., Wang, Z.W., An, H.C., Liu, X.R., Wang, D.P., 2019. An example of a hazard and risk assessment for debris flows-A case study of Niwan Gully, Wudu, China. Eng. Geol. 263 https://doi.org/10.1016/j.enggeo.2019.105351.

Qi, H.L., Yin, C., Tian, W.P., Li, J.C., 2015. Risk regionalization of highway geo-hazards in China based on ArcGIS. J. Chang'an Univ. (Nat. Sci. Ed.) 35, 22–27. https://doi.org/10.19721/j.cnki.1671-8879.2015.05.004.

Rose, S., Engel, D., Cramer, N., Cowley, W., 2010. Automatic keyword extraction from individual documents, in: Text Mining: Applications and Theory. 10.1002/9780470689646.ch1.

Sahoo, S., Dhar, A., Kar, A., 2016. Environmental vulnerability assessment using Grey Analytic Hierarchy Process based model. Environ. Impact Assess. Rev. 56, 145–154. https://doi.org/10.1016/j.eiar.2015.10.002.

Solorzano, J., Morante-Carballo, F., Montalvan-Burbano, N., Briones-Bitar, J., Carrion-Mero, P., 2022. A systematic review of the relationship between geotechnics and disasters. Sustainability 14. https://doi.org/10.3390/su141912835.

Sun, J., Southworth, J., 2013. Retrospective analysis of landscape dynamics using normalized spectral entropy. Remote Sens. Lett. 4, 1049–1056. https://doi.org/10.1080/2150704x.2013.836286.

Tan, Q.L., Bai, M.Z., Zhou, P.G., Hu, J., Qin, X.C., 2021. Geological hazard risk assessment of line landslide based on remotely sensed data and GIS. Measurement 169. https://doi.org/10.1016/j.measurement.2020.108370.

Tandel, S.S., Jamadar, A., Dudugu, S., Ieee, 2019. A survey on text mining techniques, in: 5th International Conference on Advanced Computing and Communication Systems (ICACCS). Ieee, Coimbatore, INDIA, pp. 1022–1026. 10.1109/icaccs.2019.8728547.

Wang, S.J., Che, Y.J., Ma, X.G., 2020. Integrated risk assessment of glacier lake outburst flood (GLOF) disaster over the Qinghai-Tibetan Plateau (QTP). Landslides 17, 2849–2863. https://doi.org/10.1007/s10346-020-01443-1.

Wang, Z.Y., Lam, N.S.N., 2020. Extending Getis-Ord statistics to account for local space-time autocorrelation in spatial panel data. Professional Geographer 72, 411–420. https://doi.org/10.1080/00330124.2019.1709215.

Wang, Y., Li, S.F., Ma, H.S., Jia, Y., Jiang, Y.Y., 2023. Multi-method early identification and route optimization of vulnerable geological environment hazards on mountainous highways. Geomatics and Information Science of Wuhan University, pp. 1–15.

Xia, M., Jia, K., Zhao, W.W., Liu, S.L., Wei, X.Q., Wang, B., 2021. Spatio-temporal changes of ecological vulnerability across the Qinghai-Tibetan Plateau. Ecol. Ind. 123, 11. https://doi.org/10.1016/j.ecolind.2020.107274.

Xu, Y., Qiu, X.P., Yang, X.T., Lu, X.Y., Chen, G.J., 2020. Disaster risk management models for rural relocation communities of mountainous southwestern China under the stress of geological disasters. Int. J. Disaster Risk Reduct. 50 https://doi.org/10.1016/j.ijdrr.2020.101697.

Yang, Y.G., Yu, J.Q., Fu, Y.B., Hu, J.T., 2019. Research on geological hazard risk assessment based on the cloud fuzzy clustering algorithm. J. Intell. Fuzzy Syst. 37, 4763–4770. https://doi.org/10.3233/jifs-179311.

Yin, G., Luo, J., Niu, F., Lin, Z., Liu, M., 2021. Machine learning-based thermokarst landslide susceptibility modeling across the permafrost region on the Qinghai-Tibet Plateau. Landslides 18. https://doi.org/10.1007/s10346-021-01669-7.

Yu, H., Ma, Y., Wang, L.F., Zhai, Y.S., Wang, X.Q., Ieee, 2017. A landslide intelligent detection method based on CNN and RSG_R, in: IEEE International Conference on Mechatronics and Automation (ICMA). Takamatsu, JAPAN, pp. 40–44.

Zeng, L., Zhang, K.J., Tang, X.C., Zhang, Y.X., Li, Z.W., 2018. Mid-Permian rifting in Central China: Record of geochronology, geochemistry and Sr-Nd-Hf isotopes of bimodal magmatism on NE Qinghai-Tibetan Plateau. Gondw. Res. 57, 77–89. https://doi.org/10.1016/j.gr.2017.12.013.

Zhang, L., Chen, Q., Wang, Wen, Deng, C., Zhang, S., Li, B., Wang, Wei, Cao, X., 2022. MDERank: A masked document embedding rank approach for unsupervised keyphrase extraction, in: Proceedings of the Annual Meeting of the Association for Computational Linguistics. 10.18653/v1/2022.findings-acl.34.

Zhang, T.Y., Fan, S.H., Chen, S., Li, S.P., Lu, Y.D., 2019. Geomorphic evolution and neotectonics of the Qianhe River Basin on the southwest margin of the Ordos Block, North China. J. Asian Earth Sci. 176, 184–195. https://doi.org/10.1016/j.jseaes.2019.02.020.

Zhang, H., Xu, X., Zhang, C., Fu, Z.P., Yang, H.Z., 2024. Novel method for ecosystem services assessment and analysis of road-effect zones. Transp. Res. D Transp. Environ. 127 https://doi.org/10.1016/j.trd.2024.104057.

Zhang, Q., Yu, H., Li, Z., Zhang, G., Ma, D.T., 2020. Assessing potential likelihood and impacts of landslides on transportation network vulnerability. Transp. Res. D Transp. Environ. 82 https://doi.org/10.1016/j.trd.2020.102304.

Zhang, C., Zhang, H., Zhang, M., Gong, Q.L., 2019. Novel prediction method for highway distresses in permafrost regions based on qualitative reasoning of multidimensional and multirules cloud model. Math. Probl. Eng. 2019 https://doi.org/10.1155/2019/7910752.

Zhang, C., Zhang, H., Zhao, F., Sun, J., 2019. Understanding thermal impact of roads on permafrost using normalized spectral entropy. Sustainability 11.

Zhang, H., Zhang, C., Hu, T., Zhang, M., Ren, X., Hou, L., 2020. Exploration of roadway factors and habitat quality using InVEST. Transp. Res. D Transp. Environ. 87, 102551 https://doi.org/10.1016/j.trd.2020.102551.

Zhang, J.M., Zhu, W., Cheng, Y.Q., Li, Z.H., 2021. Landslide detection in the Linzhi-Ya'an Section along the Sichuan-Tibet Railway based on InSAR and Hot Spot Analysis methods. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13183566.

Zhao, Y.G., Feng, G.C., Wang, Y.D., Wang, X.H., Wang, Y.X., Lu, H., Xu, W.B., Wang, H.Y., 2022. A new algorithm for intelligent detection of geohazards incorporating attention mechanism. Int. J. Appl. Earth Obs. Geoinf. 113 https://doi.org/10.1016/j.jag.2022.102988.

Zhu, K.X., Xu, P.H., Cao, C., Zheng, L.J., Liu, Y., Dong, X.J., 2021. Preliminary identification of geological hazards from Songpinggou to Feihong in Mao County along the Minjiang River using SBAS-InSAR technique integrated multiple spatial analysis methods. Sustainability 13. ARTN 1017 10.3390/su13031017.

Zou, L.F., Gui, W.M., 2020. Simulation and prediction of geologic hazards and the impacts on homestay buildings in scenery spots through BIM. PLoS One 15. https://doi.org/10.1371/journal.pone.0238864.