

基于机器学习的伊犁河谷黄土区泥石流易发性评估

李志, 陈宁生, 侯儒宁, 吴铭洋, 张瀛玉龙, 杜鹏

Susceptibility assessment of debris flow disaster based on machine learning models in the loess area along Yili Valley

LI Zhi, CHEN Ningsheng, HOU Runing, WU Mingyang, ZHANG Yingyulong, and DU Peng

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.202301007>

您可能感兴趣的其他文章

Articles you may be interested in

基于遥感影像多尺度分割与地质因子评价的滑坡易发性区划

Landslide susceptibility assessment based on multi-scale segmentation of remote sensing and geological factor evaluation

李文娟, 邵海 中国地质灾害与防治学报. 2021, 32(2): 94-99

基于面积高程和面积坡度积分的泥石流物质供给能力分析

The material supply ability analysis of debris flows based on area-hypsometric integral and area-gradient integral

张静, 田述军, 侯鹏鹞 中国地质灾害与防治学报. 2021, 32(4): 9-16

基于不同评价单元和灾害熵的泥石流危险性分析

Hazard analysis of debris flows based on different evaluation units and disaster entropy: A case study in Wudu section of the Bailong river basin

李小龙, 宋国虎, 向灵芝, 罗亮, 唐良琴, 沈娜, 梁梦辉 中国地质灾害与防治学报. 2021, 32(6): 107-115

基于遗传规划的泥石流多因子融合预测方法

Multi-factors fusion method of debris flow prediction based on genetic programming

翟淑花, 冒建, 南, 刘欢欢, 王云涛, 王强强, 熊春华, 王艳梅 中国地质灾害与防治学报. 2020, 31(6): 111-115

基于灰色可拓模型的洮河下游泥石流易发性评价

$\{\text{suggestArticle.titleEn}\}$

崔传峰, 王俊豪, 崔志超, 潘国营 中国地质灾害与防治学报. 2019, 30(6): 40-48

泥石流致灾因子敏感性分析——以四川都江堰龙溪河流域为例

Sensitivity analysis of debris flow to environmental factors: a case of Longxi River basin in Dujiangyan, Sichuan Province

李彩侠, 马煜, 何元勋 中国地质灾害与防治学报. 2020, 31(5): 32-39



关注微信公众号, 获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.202301007

李志, 陈宁生, 侯儒宁, 等. 基于机器学习的伊犁河谷黄土区泥石流易发性评估[J]. 中国地质灾害与防治学报, 2024, 35(3): 129-140.
LI Zhi, CHEN Ningsheng, HOU Runing, et al. Susceptibility assessment of debris flow disaster based on machine learning models in the loess area along Yili Valley[J]. The Chinese Journal of Geological Hazard and Control, 2024, 35(3): 129-140.

基于机器学习的伊犁河谷黄土区泥石流易发性评估

李志^{1,2}, 陈宁生^{2,3}, 侯儒宁^{2,4}, 吴铭洋^{2,5}, 张瀛玉龙^{1,2}, 杜鹏^{1,2}

(1. 西藏大学工学院, 西藏拉萨 850000; 2. 中国科学院、水利部成都山地灾害与环境研究所, 四川成都 610041; 3. 高原科学与可持续发展研究院, 青海西宁 810016; 4. 中国科学院大学, 北京 100049; 5. 昆明理工大学公共安全与应急管理学院, 云南昆明 650093)

摘要: 伊犁河谷地处中-哈边境, 南北疆结合带, 是丝绸之路经济带的前沿, 该区域生态环境脆弱, 泥石流灾害多发。本研究采用随机森林(RF)、支持向量机(SVM)、逻辑回归(LR)以及决策树(DT)四种机器学习模型, 模型输入为遥感判别和野外考察确定的 398 条泥石流沟以及 14 个特征参数, 计算各个评价因子权重并对泥石流易发性进行评价, 最后绘制 ROC 曲线以及计算曲线下面积(AUC)对四种机器学习的模型的准确性进行评价。研究结果表明: (1) 泥石流高易发区主要位于深切河谷地区的天山山地以及山前坡地的黄土覆盖区域; (2) 地形起伏度、多年平均降雨量、干旱指数是控制泥石流空间发育的前三个重要因素; (3) 四种模型的验证数据集 AUC 值分别为 0.938 (RF)、0.932 (SVM)、0.89 (LR)、0.879 (DT), 随机森林模型在该区域的易发性评价中具有更好的预测能力; (4) 研究区黄土的生态植被被破坏是泥石流多发的重要原因, 应该重点进行生态治理和保护, 减少水土流失, 从源头治理泥石流灾害。

关键词: 泥石流; 伊犁河谷; 黄土; 易发性; 随机森林

中图分类号: P642.23

文献标志码: A

文章编号: 1003-8035(2024)03-0129-12

Susceptibility assessment of debris flow disaster based on machine learning models in the loess area along Yili Valley

LI Zhi^{1,2}, CHEN Ningsheng^{2,3}, HOU Runing^{2,4}, WU Mingyang^{2,5}, ZHANG Yingyulong^{1,2}, DU Peng^{1,2}

(1. College of Engineering Tibet University, Lhasa, Xizang 850000, China; 2. Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Ministry of Water Resources, Chengdu, Sichuan 610041, China; 3. Academy of Plateau Science and Sustainability, Xining, Qinghai 810016, China; 4. University of Chinese Academy of Sciences, Beijing 100049, China; 5. Faculty of Public Safety and Emergency Management, Kunming University of Science and Technology, Kunming, Yunnan 650093, China)

Abstract: The Yili Valley, located on the border between China and Kazakhstan, serves as the juncture of North and South Xinjiang, and stands as a pivotal outpost on the Silk Road Economic Belt. This area possesses a fragile ecological environment and experiences frequent debris flow disasters. In this study, four machine learning models: Random Forest (RF), Logistic

收稿日期: 2023-01-10; 修订日期: 2023-03-10

投稿网址: <https://www.zgdzzyfzxb.com/>

基金项目: 第二次青藏高原科学考察项目(2019QZKK0902); 柯西河跨境水土资源管理和水灾害防控(131C11KYSB20200033); 国家自然科学基金区域联合基金项目(U20A20110)

第一作者: 李志(1999—), 男, 四川德阳人, 地质资源与地质工程专业, 硕士研究生, 研究方向为山地灾害评价与防治。

E-mail: 1245736788@qq.com

通讯作者: 陈宁生(1965—), 男, 四川成都人, 自然地理学专业, 研究员, 博导, 研究方向为山地灾害形成机理与防治。

E-mail: chennsh@imde.ac.cn

Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT)-- were employed to evaluate the debris flow susceptibility and compute the weights of evaluation factors. The models were fed inputs comprising 398 identified debris flow channels and 14 feature parameters such as fault density, topographic relief, land use, *NDVI*, multi-year average rainfall, etc obtained through remote sensing and field surveys. Also, the accuracy of the four machine learning models was evaluated by ROC curves and calculating the Area Under the Curve (*AUC*). The research results show that: (1) High debris flow susceptibility areas are mainly located in the Tianshan Mountains in the deep river valley region and the loess-covered areas in the mountain front slopes; (2) Multi-year average rainfall, drought index, and topographic relief variability are the top three influential factors controlling the spatial development of debris flows; (3) The *AUC* values for the validation datasets of the four models were 0.938 (RF), 0.932 (SVM), 0.89 (LR), 0.79 (DT), with the Random Forest model exhibiting superior predictive capability in assessing susceptibility in the region; (4) The disruption of ecological vegetation in the loess-covered region of the study area is a significant cause of frequent debris flow occurrences. Ecological governance and protection efforts should be emphasized to reduce soil erosion and effectively mitigate debris flow disasters at their source.

Keywords: debris flow; Yili Valley; loess; susceptibility; random forest

0 引言

泥石流是山地中的一种重大地质灾害,它具有短时效、突发性和高强度的特点,给山区人民的生命和财产带来巨大安全隐患。泥石流分布于全球各处山地,1950年以来,有40个国家有泥石流灾害的记录,总共造成30万人死亡和失踪。1999—2020年,全世界平均泥石流死亡和失踪的人数达到4000多人。中国地处季风区,多强烈地震带,泥石流灾害事件暴发频繁,是世界上泥石流分布及危害集中、规模最大的国家之一。据全国地质灾害通报统计,在1999—2019年间,在中国 $167 \times 10^4 \text{ km}^2$ 的泥石流灾害区域中,造成人员死亡失踪的泥石流灾害事件268次,死亡失踪4542人^[1]。

伊犁地缘区位优势突出,是古丝绸之路的北道要冲和“一带一路”的核心区,以及新疆和全国向西开放的重要商埠和国际大通道。该区域丰富的降水和广泛分布的黄土^[2]造成泥石流灾害多发,给区域的经济建设和生命财产安全带来了严重损失。如2022年霍城县G30线连霍高速果子沟路段4191 km处突发泥石流险情(<http://www.ylxw.com.cn/shehuiminsheng/2022-05-08/1052991.html>);2015年洪水、泥石流灾害造成伊犁州13个乡镇4个国营农牧场受灾,受灾人口9741人,直接经济损失1.43亿元(<https://www.chinanews.com.cn/gn/2015/05-20/7290210.shtml>);2017年4月,伊犁州伊宁县喀拉亚尔奇乡克孜勒赛沟发生黄土滑坡堵溃型泥石流灾害^[3],造成直接经济损失50余万元。然而,由于地质结构、地形和生态条件等相关的复杂控制因素,预测该区域泥石流何时发生仍然是一项挑战性任务,因此,需要重视区域泥石流的易发性分析与主控因素研究,为重

大泥石流灾害的预警和防治提供有效指导^[2-3]。

泥石流的易发性通过研究控制因子本身和发生过程中的交互作用来分析。现阶段泥石流易发性评估研究有不同侧重,一方面侧重评估方法研究,许多研究是利用不同的评估方法对易发性进行评价,例如专家打分法^[4]是早期用于评估的方法,该方法是专家经过现场勘察后根据经验建立因子和易发性的关系;针对该方法带有强烈的主观性和难以量化因子的缺点,又发展出信息量法^[5]、层次分析法和确定性系数法^[5-6]等评估,其中层次分析法对于既有定性指标也有定量指标的问题讨论的不够,进而出现了基于统计学的机器学习算法,研究表明 Logistic 回归^[7]、随机森林^[8-9]、支持向量机^[10]、神经网络^[10]等算法可以产生较为可靠的预测结果。另一方面侧重对泥石流形成主控因子的研究,包括研究泥石流地形地貌因素^[11],强干旱和强降雨^[12]等极端气候因素;岩性^[13]和强地震^[14]等地质与构造活动因子;研究物源因素可以判断泥石流是否有启动的物质基础^[15],这对评估泥石流的易发性和预测是必要的,例如土地利用和植被覆盖条件;同样,崩塌和滑坡因素也会造成泥石流物源数量增加从而影响泥石流易发程度;此外,一些研究也表明干旱河谷区的干湿循环作用会大大增加泥石流的数量与规模^[16],这与本文认为黄土与干旱影响泥石流易发程度的观点一致。目前在研究伊犁地区泥石流主控因子时仍存在急需解决的问题,第一,对控制因子贡献比值定量化的研究还较少且多将条件考虑为降雨和地貌等因素,而不考虑黄土与干旱因素。第二,仅用单一模型进行预测,预测精度无法保证,通过对比寻找最可靠的模型是解决该问题的重要手段^[17-19]。针对以上问题,为了进一步揭示控制因子对泥石流的影响,本

文基于 4 种机器学习算法, 探测不同因子的重要性程度来评估易发性。最后, 量化控制因子的贡献值以及对伊犁河谷地区进行易发性制图。

1 研究区概况

新疆伊犁谷地是中国连接中亚、东欧等国的重要枢纽, 是中国“一带一路”战略的重要节点, 位于天山造

山带(中国区域)西部(图 1)。伊犁谷地属南天山、北天山分隔而成的由断陷盆地与山间河谷间布的中亚内陆河流域, 由 3 条自西向东逐渐收缩的山脉环抱, 形似向西开口喇叭, 海拔东部较高, 西部较低^[17-18], 平均海拔超过 3 500 m。伊犁属大陆性温带半干旱气候区, 受西风带影响加上谷地独特的地形特点^[19], 使其成为新疆降雨量最多的地方, 年降水量在 250~550 mm^[20]。

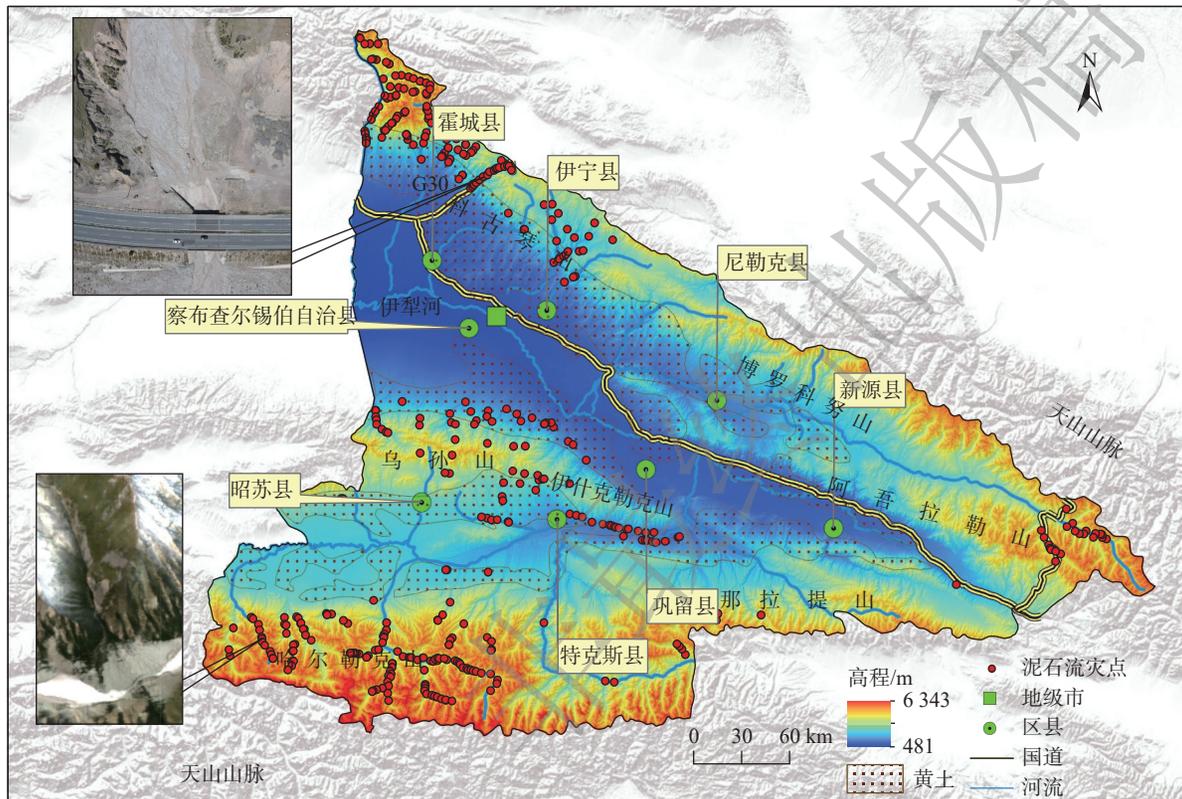


图 1 研究区概况

Fig. 1 Overview of the study area

在新生代以来, 印度洋板块与欧亚板块碰撞, 受喜马拉雅强烈造山运动影响, 伊犁盆地区域构造应力以近南北向挤压为主。据国家地震台网中心记录的历史地震数据, 自 1950 年以来, $M_S 5.0$ 级以上地震 10 次, $M_S 5.5$ 级以上地震 6 次。研究区内地层岩性复杂(图 2), 地层在垂向上具有三层结构, 可分为底部上元古界变质基底, 中部石炭系裂谷火山岩系变形基底, 上部二叠系盆地沉积岩系及厚层第四系黄土^[20-21]。黄土是由西风带带入形成广泛的风成沉积, 主要分布山前地区, 沉积厚度从几米到近一百米不等, 东部盆地比西部盆地厚。

2 数据和方法

2.1 泥石流数据

结合 Sentinel-2 和 Landsat 8 等多时相高分辨率影

像以及无人机航拍影像解译, 分别于 2020 年 9 月和 2022 年 7 月前往研究区域实地考察, 构建伊犁河谷地区泥石流灾害数据库, 库中共有 398 条泥石流, 包含有每条泥石流的流域形态、沟长、比降、泥石流类型、所属水系等信息, 其中 36% 为暴雨型泥石流、34% 为冰川型泥石流、30% 为混合型泥石流, 输入模型的正样本为 398 条泥石流流域, 负样本在使用 archydro 工具提取和人工修正的流域数据集随机选取, 虽然在流域数据集中随机选取可能会选入未被识别的真实泥石流样本, 从而增大正样本的分类结果, 但是多数的研究证明随机选择对建模是有益的。泥石流的易发性评估是典型的二分类问题, 正负样本需要一个相对平衡的比例确保模型不会偏向某种分类结果^[22]。最终选择 398 条“1”事件(有泥石流发生)与 398 条“0”事件(无泥石流发生), 将 796 条总样

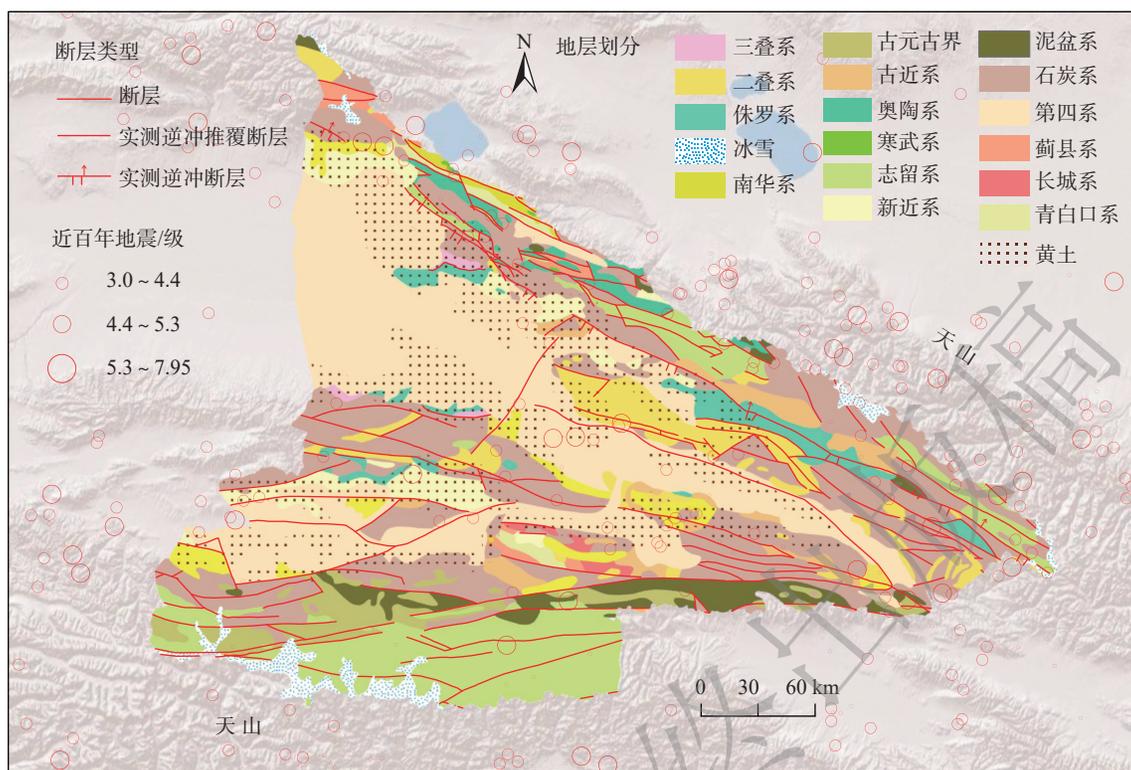


图 2 研究区地质图

Fig. 2 Geological map of the study area

本分为 70% 的训练集和 30% 的验证集。总体流程见图 3。

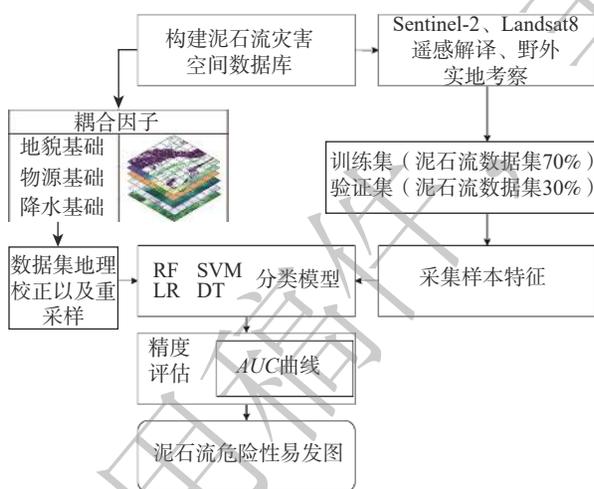


图 3 研究技术路线

Fig. 3 Research methodology and technology approach

2.2 泥石流影响因子

伊犁泥石流发生有 3 点原因：首先降雨融雪增多为泥石流提供降水基础；其次陡峭地形提供足够的势能基础，同时也决定了泥石流水文特征；最后研究区构造活动强烈，断裂发育，主要为二叠系沉积岩系和第四系黄

土，内部节理发育的岩体和遇水湿陷的黄土奠定了丰富的物源基础。本文总结前人研究的基础，考虑伊犁泥石流特征和孕灾环境以及发育情况，从地形条件、物源条件、降水条件 3 个方面出发确定 14 个影响因子(图 4)；数据来源见表 1。

(1) 地形地貌相关因子

流域面积反映流域大小基本信息，间接反应物源丰富度和降水汇集量，一般情况下流域面积越大潜在的物源与水源越多。

高差与沟床比降关系密切，为泥石流启动提供势能基础。

坡度^[11]反映物源的稳定度和径流速率，随着坡度增加，物源稳定性下降。

地形起伏度反应地表破碎程度的指标，破碎的地面起伏不平，多斜坡，导致地表岩土体稳定性降低，易形成地表径流。选择最佳提取窗口计算地形起伏度^[23-24]。计算公式为：

$$RDLS = H_{max} - H_{min} \quad (1)$$

式中： H_{max} ——最佳提取窗口内最大高程值/m；

H_{min} ——最佳提取窗口内最小高程值/m。

高程变异系数反映河流下切速率和区域隆升速率^[25]，计算公式为：

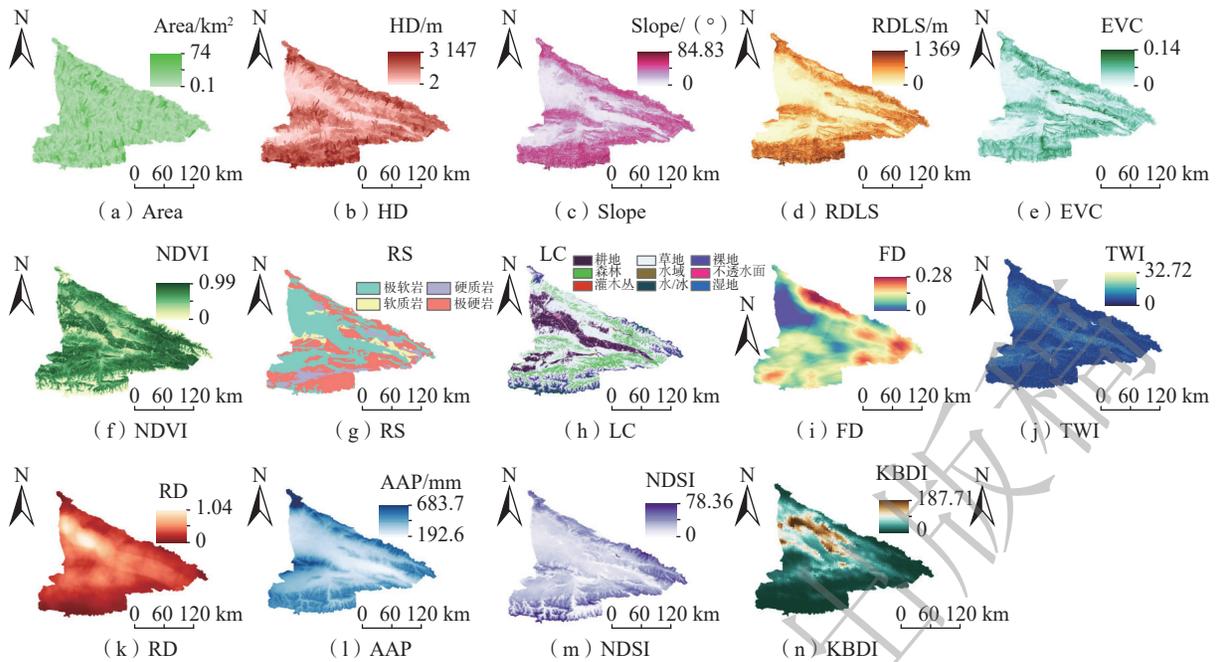


图 4 致灾因子空间分布特征

Fig. 4 Spatial distribution characteristics of disaster-causing factors

表 1 泥石流致灾因子多源异构数据来源

Table 1 Multi sources of heterogeneous data for debris flow causation factors

因子	因子(英文)	格式、分辨率/m	数据来源
流域面积	AREA	shape file	GIS分析及目视解译
高差	HD	shape file	SRTM-30 m GIS分析
坡度	Slope	Tiff/30×30	SRTM-30 m GIS分析
地形起伏度	RDLS	Tiff/30×30	SRTM-30 m GIS分析
高程变异系数	EVC	Tiff/30×30	SRTM-30 m GIS分析
归一化植被指数	NDVI	Tiff/500×500	MODIS植被指数产品(1990—2020年)(https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php)
地层岩性	RS	shape file	地质云1:20万地质图
土地利用类型	LC	Tiff/30×30	欧空局Release of Esa's Worldcover MAP(https://esa-worldcover.org/en/data-access)
断层密度	FD	shape file	地质云1:20万地质图
地形湿度指数	TWI	Tiff/30×30	SRTM-30 m GIS分析
道路密度	RD	shape file	1:25万全国基础地理数据库(https://www.webmap.cn/)
多年平均年降雨量	AAP	Tiff/30×30	Google earth engine 下载CHIRPS Daily: Climate Hazards Group InfraRed Precipitation With Station Data (Version 2.0 Final)数据集(https://earthengine.google.com/)
归一化差异积雪指数	NDSI	Tiff/500×500	MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6 (MOD10A1) (https://modis.gsfc.nasa.gov/data/dataproduct/mod13.php)
干旱指数	KBDI	Tiff/1 000×1 000	Google earth engine 下载Keetch-Byram Drought Index数据集(https://earthengine.google.com/)

$$EVC = H_{std}/H_{mean} \quad (2)$$

式中: H_{std} ——提取窗口内所有栅格的高程标准差/m;

H_{mean} ——提取窗口内所有栅格的平均高程/m。

(2) 物源相关因子

归一化植被指数反映地表植被生长状态,是表征区域地表植被覆盖的最佳指示因子。植被是影响地质灾害的重要因子,在植被指数高的区域内,受风化作用小,降雨后不易快速形成地表径流。计算公式为:

$$NDVI = (NIR - R)/(NIR + R) \quad (3)$$

式中: NIR ——近红外波段的反射值;

R ——红光波段的反射值。

地层岩性^[13]反映了泥石流的主要物源类型,风化作用导致岩体破碎,破碎岩体进入沟道内,对泥石流物源进行补给。将研究区的岩性根据工程地质岩组分类极软岩、软质岩、硬质岩、极硬岩4种岩性。

土地利用类型反映了区域内地表覆盖物类型和

松散物数量,人类对土地利用类型有干预作用,如对生态环境的破坏会导致和加剧泥石流灾害的产生。不同土地利用类型的地表渗透速度和抗冲刷能力有较大差异。

断层密度反映了岩体裂隙发育程度,岩体裂隙发育导致雨水下渗,促进斜坡失稳,对泥石流物源进行补给。

地形湿度指数反映了土壤内水分的干湿程度,土壤含水量越大越易达到饱和状态,从而也更容易产生径流,反之则土壤越干。计算公式为:

$$TWI = \ln\left(\frac{S}{\tan\beta}\right) \quad (4)$$

式中: S ——单位等高线长度上的区域面积/($m^2 \cdot m^{-1}$);

$\tan\beta$ ——区域地形坡度的正切值。

道路密度反映了人类对自然环境的干扰强度,人类活动下能够快速改造自然,特别是道路工程项目下,人类可大规模改造地质地貌对稳定土体产生扰动。

干旱指数反映地区干旱程度,长历时的干旱缺水会导致坡体地下水位线降低,土体含水量和孔隙水压力值降低,导致土体结构变松散和结构强度降低。

(3) 降水相关因子

研究表明降水是诱发泥石流的重要动力因素,特别是在“厄尔尼诺”现象下的极端降水更易作为泥石流启动条件^[26]。对于处在干旱气候的区域,主要降雨

发生于夏季,多年平均年降雨量基本反映了区域内降雨强度。

归一化差异积雪指数反映了区域内积雪程度,伊犁流域内存在冰川的地区,积雪也是降水的主要来源,容易形成冰川泥石流。计算公式为:

$$NDSI = (Green - SWIR1)/(Green + SWIR1) \quad (5)$$

式中: $Green$ ——绿色波段的反射值;

$SWIR1$ ——短波红外波段的反射值。

初始空段落

2.3 数据预处理

数据的预处理分为数据清洗和数据处理以及检测多重共线性。收集的数据集通常是含有缺失值,为了在不完备数据集上构建有效的机器学习模型,需要对数据集进行清洗。数据处理将非栅格数据转化为栅格数据,并在同一投影坐标系下将因子重采样至相同分辨率(像元大小 30×30);机器学习算法输入数据为文本数据,按照流域单元统计因子数值作为建模数据。因子之间的共线性会极大破坏回归拟合过程以及影响模型准确性。因此在数据输入之前需要对前文选取的 14 个因子进行共线性检验,通过皮尔逊(pearson)相关系数计算因子之间的相关性大小(图 5),根据相关矩阵的热力图去除高差(pearson>8)这个因子,选择剩下的 13 个因子进行建模。

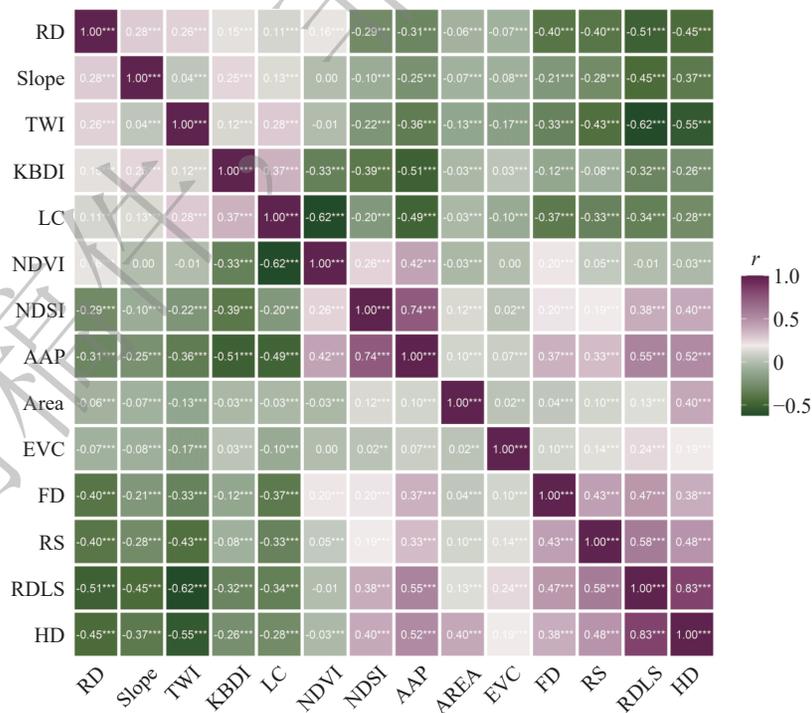


图 5 致灾因子 Person 相关系数热力图

Fig. 5 Heatmap of person correlation coefficients of the causative factor

2.4 方法

随机森林(RF)是由 Breiman 基于集成思想提出是由多个决策树构成的集成学习模型。作为数据驱动的非参数分类方法, RF 使用 CART 集群进行预测分类。利用随机抽样技术通过有放回地抽取训练样本集的一部分创建回归树集群, 通过投票方式得到最终分类结果^[27]。通过随机分割将样本集分为不同的训练集与测试集, 约三分之二的训练样本(袋内样本)会用于训练回归树, 剩余三分之一样本(袋外样本)用于内部验证, 以估计随机森林分类器分类结果的误差, 称袋外误差。

支持向量机(SVM)是一种二分类监督分类器, 基于结构风险最小化理论和统计学原理 VC 维理论对数据进行分类的方法, 通过映射函数将原有的样本数据投影到更高维度的特征空间上, 寻找超平面将数据尽可能多的正确分开, 从而实现在原空间中更好的分类效果^[28]。

逻辑回归模型(LR)常用于二分类问题, 主要是在一个因变量和多个自变量之间形成多元回归关系, 其中自变量可以是离散变量或者连续变量, 模型本质是假设数据服从这个分布, 然后使用极大似然估计做参数的估计。

决策树是一种树形结构, 树以代表训练样本的单个结点开始, 如果样本都在同一个类, 则该结点成为树叶, 并用该类标记。否则, 将选择最有分类能力的属性作为决策树的当前结点。根据当前决策结点属性取值的不同, 一旦一个属性出现在一个结点上, 就不必在该结点的任何后代考虑它。这说明其中每个内部节点表示一个属性上的判断, 每个分支代表一个判断结果的输出, 最后每个叶节点代表一种分类结果。

3 结果

3.1 易发性分区及制图

基于自然间断法将伊犁河谷流域四种模型的易发性划分为极低、低、中等、高、极高五类易发等级(图 6)。极高和高易发分区主要集中在科古琴山和博罗科努山以及阿吾拉勒山的北区、乌孙山和伊什克勒克山的中部区域、哈尔勒克山和那拉提山构成的南部区域。整体而言, 4 种模型的易发性分区相似度较高, 从地形地貌形态来看极高和高易发性区域均主要分布在侵蚀风化剧烈、气候干旱的山前地带以及山地区域, 占总面积在 41.9% 到 45.4%(表 2), 山前地带由于黄土

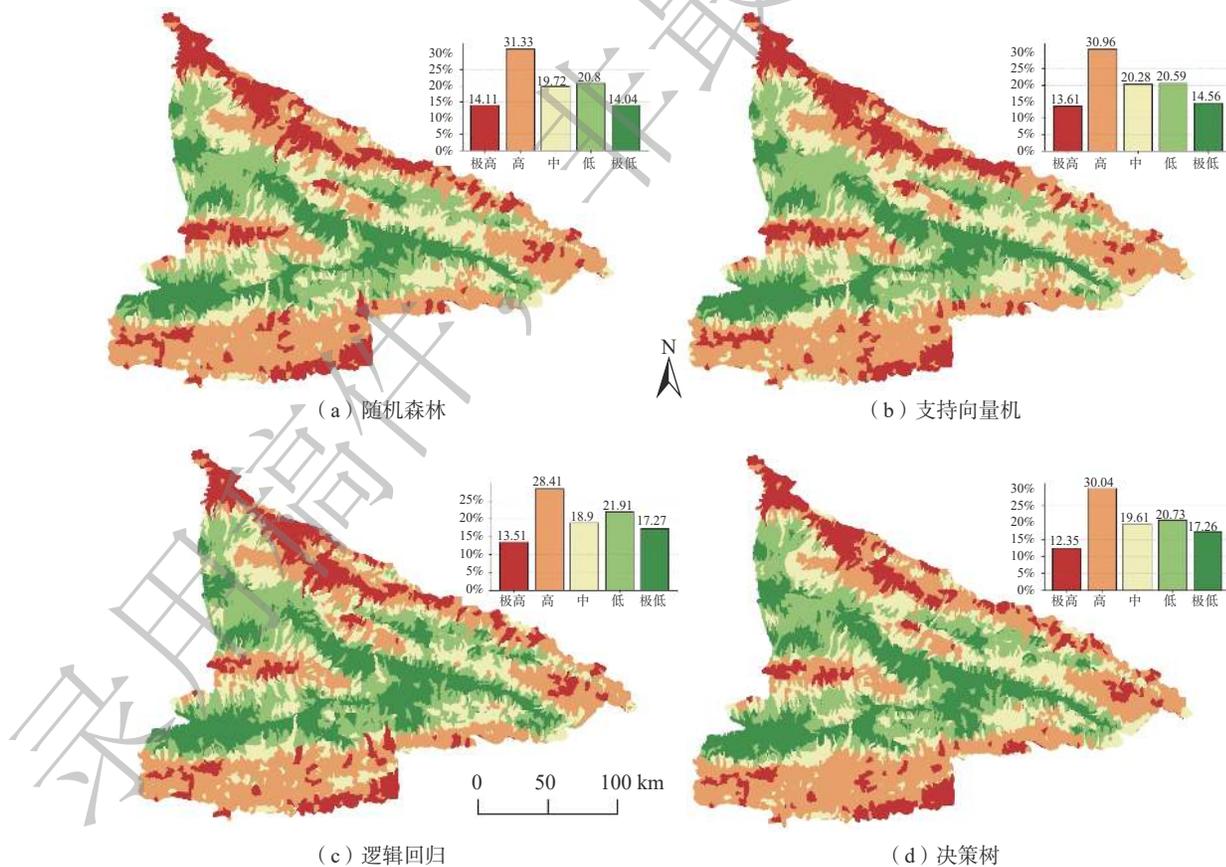


图 6 四种模型易发性分级

Fig. 6 Susceptibility classification using four models

表 2 四种模型易发性分区统计

Table 2 Statistical analysis of susceptibility zoning using four models

模型	易发性分区	面积/km ²	面积占比/%	分区内泥石流条数/条	泥石流条数占比/%
随机森林	极高	7 971.642	14.11	143	35.93
	高	17 701.790	31.33	177	44.47
	中	11 144.620	19.72	46	11.56
	低	11 751.190	20.80	23	5.78
	极低	7 930.759	14.04	9	2.26
支持向量机	极高	7 689.020	13.61	152	38.19
	高	17 493.350	30.96	169	42.46
	中	11 457.150	20.28	45	11.31
	低	11 635.640	20.59	23	5.78
	极低	8 224.840	14.56	9	2.26
逻辑回归	极高	7 634.988	13.51	127	31.91
	高	16 051.380	28.41	178	44.72
	中	10 676.710	18.90	50	12.56
	低	12 377.990	21.91	28	7.04
	极低	9 758.932	17.27	15	3.77
决策树	极高	6 979.475	12.35	129	32.41
	高	16 973.330	30.04	175	43.97
	中	11 080.390	19.61	56	14.07
	低	11 712.370	20.73	26	6.53
	极低	9 754.436	17.26	12	3.02

堆积区的特殊性导致容易发生暴雨型泥石流,如霍城县南部、伊宁县、伊宁市等低山丘陵地带,而昭苏县南部、特克斯县南部、新源县南部和霍城县北部等山地同样也是泥石流易发区域。随机森林模型的极高和高易发性区域主要位于科古琴山、哈尔勒克山的山地以及山前风成黄土地带,面积约 25 673.4 km²,占研究区总面积的 45.4%。

3.2 特征重要性

地形起伏度、雨季降雨、干旱度是伊犁地区泥石流发育分布的主要控制因素,人类活动对区域泥石流发育没有显著控制作用。4种模型对伊犁河谷流域 13 个控制因子的相对重要性结果见图 7,根据建模结果发现随机森林模型对 13 个控制因子对泥石流发育贡献值由大到小依次为地形起伏度(25.49%)、多年平均降雨

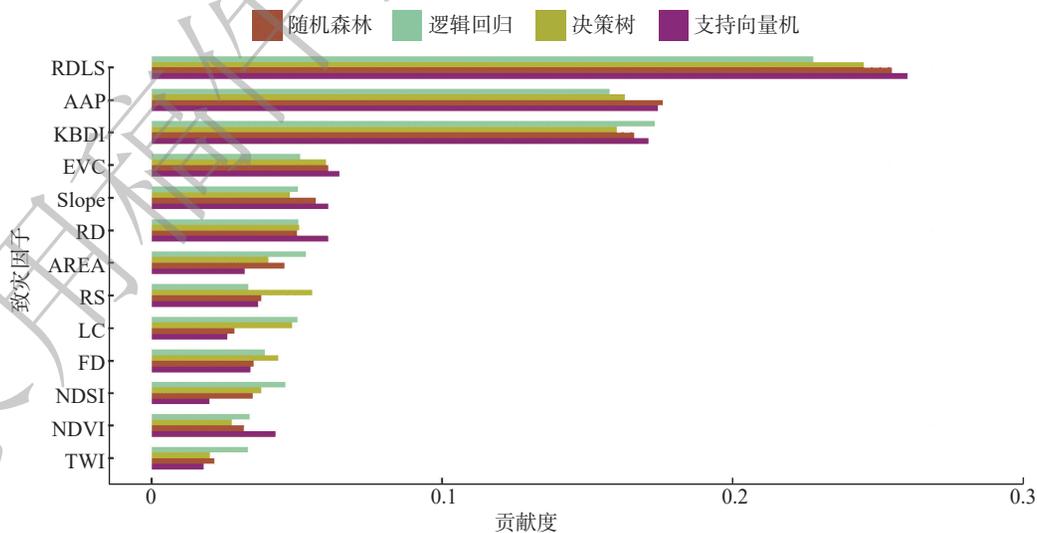


图 7 不同模型致灾因子贡献重要性

Fig. 7 Importance of causative factors contribution in different models

量(17.58%)、干旱指数(16.61%)、坡度(5.66%)、道路距离(5.02%)、流域面积(4.6%)、地层岩性(3.77%)、断层密度(3.51%)、归一化积雪指数(3.49%)、归一化植被指数(3.18%)、土地利用类型(2.86%)、地形湿度指数(2.17%)。

3.3 结果评估以及验证

为证明评价模型的可靠性,采用已被广泛用于分类模型性能评估的受试者工作特征曲线(ROC)^[29]。ROC曲线的横轴代表非灾害点预测为灾害点的比例即假阳性率,纵坐标表示灾害点预测准确性的占比即真阳性率,曲线下方的面积代表模型的准确性,曲线下面积(AUC)值越接近1,性能越好。在对样本的预测结果进行计数后得到ROC曲线,结果如图8所示,其中支持向量机、逻辑回归、决策树模型的AUC值大于0.8,说明模型结果具有较好的可信度,而随机森林模型的AUC值(0.936)是4种模型中可信度最高的模型。将随机森林模型评估结果与已查明的泥石流沟对比分析,结果表明:已查明的398条泥石流沟中,35.93%位于极高易发性分区,44.47%位于高易发性分区内,高于其他3种模型占比,这表明随机森林的模型相比其余三种模型在识别泥石流灾害方面具有最强的空间识别能力。我们选择可信度最高的随机森林模型与野外调查(图9中A、B、C为野外验证区域)得到的灾害事件进行验证(图9),2020年和2022年的夏季野外调查中,发现霍城县G30道路果子沟段发生了超过10余处泥石流灾害事件(图10);典型案例如阿力马力萨沟(图10A),根据走访调查,当地牧民表示不但2020年与2022年雨季期间多次暴发泥石流,而且在历史上也曾多次暴发,该沟沟道及岸坡侵蚀现象十分严重,属于高频次活动的泥石流。

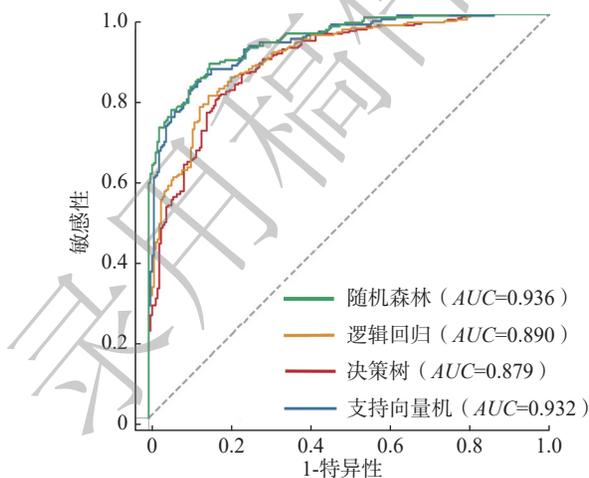


图8 不同模型 ROC 曲线和 AUC 对比

Fig. 8 Comparison of ROC curves and AUC among different models

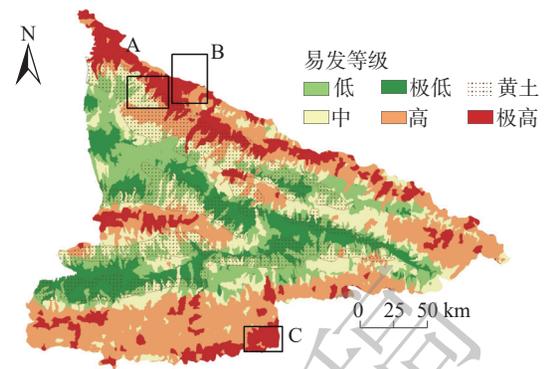


图9 随机森林模型野外验证区域

Fig. 9 Field validation area for the random forest model

4 讨论

在本研究中,利用随机森林模型制作泥石流易发图,结合黄土分布数据发现流域内泥石流易发区域与黄土分布存在关联,不考虑平原区域黄土,山前丘陵区70%的高易发区域位于黄土范围之内。这可能是黄土自身孔隙度高、黏土含量高、易发生水软等特性加剧引起松散土体斜坡不稳定性造成,从而为泥石流运动提供了丰富的物源。根据野外考察,该区域多见黄土的坡脚侵蚀和坡体植被退化现象(图10B),这跟区域日益增长的人口和放牧活动相关。黄土区地表植被以草地为主,一旦被破坏,较少的区域降水限制了生态植被恢复速率,所以治理区域泥石流灾害工作,从源头保护区域生态、减少水土流失十分必要。

对于因子特征重要性而言,地形起伏度贡献度是13个因子中分值最大的,极高易发区内的平均地形起伏度为1000 m,这为泥石流发育提供了较大的重力势能。与其他区域相比,伊犁地区是典型的河谷地貌,地形起伏度较其他泥石流广泛发育的龙门山、横断山山区平缓,并非所有区域都达到了泥石流形成的地形阈值条件,所以地形因素是该区域泥石流发育的第一控制因素。雨季降雨量和干旱指数分别是分值第二、第三的因子,强调了干旱和雨季极端降雨对泥石流发育形成的重要控制作用:一方面干旱造成土体坡面干裂,为降雨在坡体土壤的优势渗流提供通道^[30-31],另一方面,干旱会造成坡体表层土壤“气阻”效应的出现,降雨会造成表层土壤的快速饱和与“超渗产流”。由以上两点原因,极端干旱后的极端降雨会触发灾害性泥石流。另外,道路密度揭示了人类对于自然环境的扰动程度,该因素对区域泥石流发育的控制程度相对较小,这与中国西南区域、尤其是东南闽浙山地地区的研究结果相反,这可能归因于该区域较少的人口密度和极低的人类活



图 10 野外考察验证图集

Fig. 10 Field investigation verification gallery

动强度。

本研究不足之处如下。第一，泥石流数据清单是根据遥感解译和野外调查得到的，但对于某些泥石流沟还无法确定其发生年份，未来应在时间尺度上探究泥石流发育规律。第二，对于评价单元的选择，本研究的流域选择尺度未达到较小流域尺度，也许在已选择流域的子流域会发生泥石流，而其他子流域不会发生泥石流，这被称为空间尺度效应，尽管一些研究已经比较了不同评价单元对于易发性评估的影响，但对泥石流灾害易发性评估的最好评价单元还没有定论，值得肯定的是在本文的评价单元下，评估精度可以接受。这也是未来研究中需要考虑的一个重要问题。

5 结论

该研究主要对伊犁河谷流域的泥石流地质灾害进行易发性制图。考虑了 14 个条件因素，并计算了它们

的 person 相关性，以验证每个变量不存在多重共线性。RF、LR、DT 和 SVM 4 个模型用于绘制泥石流易发性图，此外，对模型结果进行了分析，确定泥石流灾害与各类条件因素之间的定量关系。研究结果有助于了解研究区易发性情况以及加深对泥石流控制因素的认识，对制定泥石流减灾措施具有重要的参考价值。本研究的主要结论可归纳如下：

(1) 研究区极高易发面积 7 971.64 km²，占总面积的 14.11%；高易发面积 17 701.8 km²，占总面积的 31.33% 伊犁河谷流域中 76% 的已查明泥石流发育在深切河谷山地区域，24% 分布在山前黄土覆盖地带。泥石流高易发区域主要集中在霍城县、伊宁县、新源县南部、昭苏县南部、特克斯县南部。

(2) 地形起伏度、雨季降雨、干旱度是控制伊犁地区泥石流发育分布的前三个重要因子，地形明显控制着伊犁地区泥石流发育情况，相反人类活动没有显著

控制作用。

(3)对于模型的预测准确性,根据 ROC 曲线,随机森林、支持向量机、逻辑回归、决策树模型的 AUC 值分别为是 0.936、0.932、0.89 和 0.879,这表明随机森林模型在区域尺度的易发性评价中具有更为优秀的预测能力。

(4)研究区黄土的生态植被被破坏是泥石流多发的重要原因,应该重点进行生态治理和保护,减少水土流失,从源头治理泥石流灾害。

参考文献(References):

- [1] 陈宁生,田树峰,张勇,等.泥石流灾害的物源控制与高性能减灾 [J] . *地学前缘*, 2021, 28(4): 337-348. [CHEN Ningsheng, TIAN Shufeng, ZHANG Yong, et al. Soil mass domination in debris-flow disasters and strategy for hazard mitigation [J] . *Earth Science Frontiers*, 2021, 28(4): 337-348. (in Chinese with English abstract)]
- [2] 宋友桂,史正涛.伊犁盆地黄土分布与组成特征 [J] . *地理科学*, 2010, 30(2): 267-272. [SONG Yougui, SHI Zhengtao. Distribution and compositions of loess sediments in Yili Basin, central Asia [J] . *Scientia Geographica Sinica*, 2010, 30(2): 267-272. (in Chinese with English abstract)]
- [3] 邵海,魏云杰,黄喆,等.新疆伊宁克孜勒赛黄土滑坡堵溃型泥石流成灾模式 [J] . *中国地质灾害与防治学报*, 2018, 29(6): 40-46. [SHAO Hai, WEI Yunjie, HUANG Zhe, et al. Kezilesai loess landslide dam-breaking debris flow hazards model in Yining County, Xinjiang [J] . *The Chinese Journal of Geological Hazard and Control*, 2018, 29(6): 40-46. (in Chinese with English abstract)]
- [4] AHMED B, DEWAN A. Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in Chittagong City corporation, Bangladesh [J] . *Remote Sensing*, 2017, 9(4): 304.
- [5] LI Yongchao, CHEN Jianping, TAN Chun, et al. Application of the borderline-SMOTE method in susceptibility assessments of debris flows in Pinggu District, Beijing, China [J] . *Natural Hazards*, 2021, 105(3): 2499-2522.
- [6] 徐艳琴,白淑英,徐永明.基于两种方法的攀西泥石流易发性评价对比分析 [J] . *水土保持研究*, 2018, 25(3): 285-291. [XU Yanqin, BAI Shuying, XU Yongming. Comparative analysis of debris flow susceptibility assessment based on two methods in Panxi District [J] . *Research of Soil and Water Conservation*, 2018, 25(3): 285-291. (in Chinese with English abstract)]
- [7] LIANG Zhu, WANG Changming, ZHANG Zhimin, et al. A comparison of statistical and machine learning methods for debris flow susceptibility mapping [J] . *Stochastic Environmental Research and Risk Assessment*, 2020, 34(11): 1887-1907.
- [8] 黄发明,胡松雁,闫学涯,等.基于机器学习的滑坡易发性预测建模及其主控因子识别 [J] . *地质科技通报*, 2022, 41(2): 79-90. [HUANG Faming, HU Songyan, YAN Xueya, et al. Landslide susceptibility prediction and identification of its main environmental factors based on machine learning models [J] . *Bulletin of Geological Science and Technology*, 2022, 41(2): 79-90. (in Chinese with English abstract)]
- [9] 刘永焱,第宝锋,詹宇,等.基于随机森林模型的泥石流易发性评价——以汶川地震重灾区为例 [J] . *山地学报*, 2018, 36(5): 765-773. [LIU Yongyao, DI Baofeng, ZHAN Yu, et al. Debris flows susceptibility assessment in Wenchuan earthquake areas based on random forest algorithm model [J] . *Mountain Research*, 2018, 36(5): 765-773. (in Chinese with English abstract)]
- [10] DI Baofeng, ZHANG Hanyue, LIU Yongyao, et al. Assessing susceptibility of debris flow in southwest China using gradient boosting machine [J] . *Scientific Reports*, 2019, 9: 12532.
- [11] QING Feng, ZHAO Yan, MENG Xingmin, et al. Application of machine learning to debris flow susceptibility mapping along the china-pakistan karakoram highway [J] . *Remote Sensing*, 2020, 12(18): 2933.
- [12] MARINO P, SIVA SUBRAMANIAN S, FAN Xuanmei, et al. Changes in debris-flow susceptibility after the Wenchuan earthquake revealed by meteorological and hydro-meteorological thresholds [J] . *CATENA*, 2022, 210: 105929.
- [13] ESPER ANGILLIERI M Y. Debris flow susceptibility mapping using frequency ratio and seed cells, in a portion of a mountain international route, dry central andes of argentina [J] . *CATENA*, 2020, 189: 104504.
- [14] 李郎平,兰恒星,郭长宝,等.基于改进频率比法的川藏铁路沿线及邻区地质灾害易发性分区评价 [J] . *现代地质*, 2017, 31(5): 911-929. [LI Langping, LAN Hengxing, GUO Changbao, et al. Geohazard susceptibility assessment along the sichuan-tibet railway and its adjacent area using an improved frequency ratio method [J] . *Geoscience*, 2017, 31(5): 911-929. (in Chinese with English abstract)]
- [15] WANG Jun, YU Yan, YANG Shun, et al. A modified certainty coefficient method (M-CF) for debris flow susceptibility assessment: A case study for the Wenchuan earthquake meizoseismal areas [J] . *Journal of Mountain Science*, 2014, 11(5): 1286-1297.
- [16] 陈宁生,邓明枫,胡桂胜,等.地震影响下西南干旱山区泥石流危险性特征与防治对策 [J] . *四川大学学报(工程科学版)*, 2010, 42(增刊1): 1-6. [CHEN Ningsheng, DENG Mingfeng, HU Guisheng, et al. Risk characteristics and prevention strategy of debris flow under the seismic influence in mountainous arid area, southwest China [J] . *Journal of Sichuan University (Engineering Science Edition)*, 2010, 42(Sup 1): 1-6. (in Chinese with English abstract)]
- [17] SONG Yougui, SHI Zhengtao, FANG Xiaomin, et al. Loess magnetic properties in the Ili Basin and their correlation with the

- Chinese Loess Plateau [J] . *Science China Earth Sciences*, 2010, 53(3): 419 - 431.
- [18] 张军民. 伊犁河流域地质构造及其地形地貌特点的研究 [J] . *石河子大学学报(自然科学版)*, 2006, 24(4): 442 - 445. [ZHANG Junmin. Studies on the geological structures and characteristic of terrain and landform in Yili River Basin [J] . *Journal of Shihezi University (Natural Science)*, 2006, 24(4): 442 - 445. (in Chinese with English abstract)]
- [19] 屈文军, 张小曳, 王丹, 等. 西风带研究的重要意义 [J] . *海洋地质与第四纪地质*, 2004, 24(1): 125 - 132. [QU Wenjun, ZHANG Xiaoye, WANG Dan, et al. The important significance of westerly wind study [J] . *Marine Geology & Quaternary Geology*, 2004, 24(1): 125 - 132. (in Chinese with English abstract)]
- [20] SUN Huilan, CHEN Yaning, LI Weihong, et al. Variation and abrupt change of climate in Ili River Basin, Xinjiang [J] . *Journal of Geographical Sciences*, 2010, 20(5): 652 - 666.
- [21] 冯建辉, 陶国强, 梅志超, 等. 新疆伊犁盆地地层划分与对比 [J] . *断块油气田*, 1996, 3(3): 22 - 28. [FENG Jianhu, TAO Guoqiang, MEI Zhichao, et al. Stratigraphic classification and correlation in Yili Basin, Xinjiang [J] . *Fault-Block Oil & Gas Field*, 1996, 3(3): 22 - 28. (in Chinese with English abstract)]
- [22] KORNEJADY A, OWNEGH M, BAHREMAND A. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods [J] . *CATENA*, 2017, 152: 144 - 162.
- [23] 张明媚, 薛永安. 斜坡地质灾害敏感性评价中地势起伏度提取最佳尺度研究 [J] . *太原理工大学学报*, 2020, 51(6): 881 - 888. [ZHANG Mingmei, XUE Yongan. Optimal scale for extracting relief amplitude in slope geological hazard sensitivity evaluation [J] . *Journal of Taiyuan University of Technology*, 2020, 51(6): 881 - 888. (in Chinese with English abstract)]
- [24] 张伟, 李爱农. 基于 DEM 的中国地形起伏度适宜计算尺度研究 [J] . *地理与地理信息科学*, 2012, 28(4): 8 - 12. [ZHANG Wei, LI Aiong. Study on the optimal scale for calculating the relief amplitude in China based on DEM [J] . *Geography and Geo-Information Science*, 2012, 28(4): 8 - 12. (in Chinese with English abstract)]
- [25] 杨晓平, 王萍, 李晓峰, 等. 地形坡度和高程变异系数在识别墨脱活动断裂带中的应用 [J] . *地震地质*, 2019, 41(2): 419 - 435. [YANG Xiaoping, WANG Ping, LI Xiaofeng, et al. Application of topographic slope and elevation variation coefficient in identifying the Motuo active fault zone [J] . *Seismology and Geology*, 2019, 41(2): 419 - 435. (in Chinese with English abstract)]
- [26] 高云建, 陈宁生, 胡桂胜, 等. 西南山区泥石流灾害与厄尔尼诺-拉尼娜事件时空耦合关系分析 [J] . *长江科学院院报*, 2019, 36(4): 43 - 48. [GAO Yunjian, CHEN Ningsheng, HU Guisheng, et al. Temporal and spatial coupling relationship between debris flow and El Nino-La nina event in southwest China [J] . *Journal of Yangtze River Scientific Research Institute*, 2019, 36(4): 43 - 48. (in Chinese with English abstract)]
- [27] 于新洋, 赵庚星, 常春艳, 等. 随机森林遥感信息提取研究进展及应用展望 [J] . *遥感信息*, 2019, 34(2): 8 - 14. [YU Xinyang, ZHAO Gengxing, CHANG Chunyan, et al. Random forest classifier in remote sensing information extraction: A review of applications and future development [J] . *Remote Sensing Information*, 2019, 34(2): 8 - 14. (in Chinese with English abstract)]
- [28] 黄发明, 殷坤龙, 蒋水华, 等. 基于聚类分析和支持向量机的滑坡易发性评价 [J] . *岩石力学与工程学报*, 2018, 37(1): 156 - 167. [HUANG Faming, YIN Kunlong, JIANG Shuihua, et al. Landslide susceptibility assessment based on clustering analysis and support vector machine [J] . *Chinese Journal of Rock Mechanics and Engineering*, 2018, 37(1): 156 - 167. (in Chinese with English abstract)]
- [29] PROVOST F J, FAWCETT T. Robust classification for imprecise environments [J] . *Machine Learning*, 2001, 42(3): 203 - 231.
- [30] 王凤娘. 2010 年极端干湿循环对我国西南山区大规模泥石流滑坡灾害的促进作用 [C] //2015 年全国工程地质学术年会论文集. 长春, 2015: 73 - 79. [WANG fengniang. Promoting impact of extreme dry-wet cycle in 2010 on the large-scale debris flows and landslides in southwest mountain areas of china [C] //Proceedings of the 2015 National Annual Conference on Engineering Geology. Changchun, 2015: 73 - 79. (in Chinese with English abstract)]
- [31] BREIMAN L. Random forests [J] . *Machine Learning*, 2001, 45(1): 5 - 32.