

DOI:10.13870/j.cnki.stbcbx.2026.03.037 CSTR:32310.14.stbcbx.2026.03.037

郭飞,黄皋羽,赖鹏,等.不同机器学习模型和输入变量的崩岗易发性评价适宜性研究[J].水土保持学报,2026,40(3):47-57.

GUO Fei, HUANG Gaoyu, LAI Peng, et al. Study on the suitability of different machine learning models and input variables for Benggang susceptibility assessment[J]. Journal of Soil and Water Conservation, 2026, 40(3):47-57.

不同机器学习模型和输入变量的崩岗易发性评价适宜性研究

郭飞^{1,2}, 黄皋羽^{1,2}, 赖鹏^{1,2}, 杨亚会³,

刘琦良⁴, 黄涛⁵, 程冬兵⁶, 陈勇^{1,2}

(1.三峡库区地质灾害教育部重点实验室,湖北宜昌443002; 2.三峡大学土木与建筑学院,湖北宜昌443002;
3.三峡大学水利与环境学院,湖北宜昌443002; 4.湖北文理学院理工学院,湖北襄阳441025; 5.湖北工程学院
新技术学院,湖北孝感432000; 6.长江水利委员会汉江流域治理保护中心,武汉430010)

摘要: [目的] 为探究不同机器学习模型和输入变量对崩岗易发性评价的适宜性。[方法] 以江西省石城县为研究区,通过地理探测器(GD)筛选构建指标体系,分别将原始值、频率比(FR)和邻域频率比(NFR)作为输入变量,对比多层感知机(MLP)和随机森林(RF)模型,开展不同评价模型与输入变量的适宜性研究。[结果] 1)NFR输入变量下MLP模型、RF模型易发性结果的AUC分别为0.854、0.860,评估精度均良好,NFR是适宜的输入变量;2)RF模型整体优于MLP模型,其中原始值-RF、NFR值-RF和FR值-RF模型的高易发区崩岗密度分别为3.93、3.83和3.69,原始值-RF对极高、高易发区识别能力最强;3)极高易发区内的崩岗密度最大,高易发区、极高易发区集中于西北部,与实际崩岗分布格局吻合较好。[结论] NFR是一种普适性较强的输入变量,与原始值和FR相比,NFR在MLP和RF模型中的稳健性最高。RF模型比MLP模型更适宜开展崩岗易发性评价。

关键词: 频率比值; 邻域频率比值; 随机森林; 多层感知机; 崩岗易发性

中图分类号:P694 文献标识码:A 文章编号:1009-2242(2026)03-0047-11

Study on the Suitability of Different Machine Learning Models and Input Variables for Benggang Susceptibility Assessment

GUO Fei^{1,2}, HUANG Gaoyu^{1,2}, LAI Peng^{1,2}, YANG Yahui³,

LIU Qiliang⁴, HUANG Tao⁵, CHENG Dongbing⁶, CHEN Yong^{1,2}

(1.Key Laboratory of Geological Hazards on Three Gorges Reservoir Area, Ministry of Education, Yichang, Hubei 443002, China; 2.College of Civil Engineering and Architecture, China Three Gorges University, Yichang, Hubei 443002, China; 3.College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang, Hubei 443002, China; 4.Science and Technology College, Hubei University of Arts and Science, Xiangyang, Hubei 441025, China; 5.College of Technology, Hubei Engineering University, Xiaogan, Hubei 432000, China; 6.Hanjiang River Basin Management and Protection Center, Changjiang Water Resources Commission, Wuhan 430010, China)

Abstract: [Objective] To investigate the suitability of different machine learning models and input variables for assessing collapsing gully susceptibility. [Methods] Taking Shicheng County, Jiangxi Province as the study area, an indicator system was constructed using geodetector (GD) for factor screening. The original values, frequency ratio (FR), and neighborhood frequency ratio (NFR) were used as input variables for the multilayer perceptron (MLP) and random forest (RF) models. The adaptability of these different models and input variables for Benggang susceptibility assessment was studied. [Results] 1) The AUC values of susceptibility assessment results from MLP and RF models under the NFR input variables were 0.854 and 0.860, respectively. Both models

收稿日期:2026-01-08 修回日期:2026-02-06 录用日期:2026-02-21 网络首发日期(www.cnki.net):2026-03-18

资助项目:国家自然科学基金项目(42107489)

第一作者:郭飞(1987—),男,博士,副教授,主要从事崩岗灾害风险评估研究。E-mail:ybbnui.2008@163.com

通信作者:杨亚会(1990—),女,博士,讲师,主要从事工程结构安全与防护研究。E-mail:yangyahui@163.com

http://stbcbx.alljournal.com.cn

demonstrated good assessment performance, indicating that NFR was a suitable input variable. 2) The RF model generally outperformed the MLP model. Specifically, the Benggang densities in high susceptibility areas of original value-RF, NFR-RF, and FR-RF models were 3.93, 3.83, and 3.69, respectively. The original value-RF model demonstrated the strongest capability in identifying extremely high and high susceptibility areas. 3) The Benggang density was highest in the extremely high susceptibility area. Both high and extremely high susceptibility areas were concentrated in the northwest, closely matching the actual distribution pattern of Benggang. [Conclusion] NFR is a highly generalizable input variable. Compared with original values and FR, NFR exhibits the highest robustness in both MLP and RF models. The RF model is more suitable than the MLP model for assessing Benggang susceptibility.

Keywords: frequency ratio; neighborhood frequency ratio; random forest; multilayer perceptron; Benggang susceptibility

Received: 2026-01-08

Revised: 2026-02-06

Accepted: 2026-02-21

Online(www.cnki.net): 2026-03-18

崩岗是一种在岩石风化壳或山坡土体受水力与重力共同作用下,经历解体、崩塌及堆积过程而形成的围椅状地貌形态^[1],是我国南方热带及亚热带地区侵蚀强度最大,危害最为严重的一种侵蚀类型,被喻为“生态溃疡”^[2]。崩岗集中分布于长江以南的广东、江西、福建、湖北、广西、湖南、安徽等7个省(自治区),且多发生于农业生产区,影响耕地面积高达38.04万hm²,导致区域粮食产量显著减少,土壤退化严重,农田恢复和整理的费用高达5.5亿元^[3-4]。崩岗侵蚀不仅直接破坏土地资源结构与功能,还对坡面耕地造成直接侵蚀。同时其产生的大量泥沙向下游输移,对下游基本农田造成严重的淤积与掩埋灾害,进而对区域社会经济与生态环境产生显著负面影响。崩岗侵蚀已成为我国南方红壤区威胁生态安全、粮食生产、防洪体系及居民安全的重要地质灾害之一^[5]。因此,建立精准的潜在崩岗预测模型,进行崩岗易发性分析,对防灾减灾和城镇区划至关重要^[6]。

输入变量方法的选取与构建是崩岗易发性评价的核心环节,直接影响模型的预测精度。输入变量的形式决定评价指标的数据表征方式,其构建决定指标数据在评价模型中的具体表达,对评价模型造成直接影响。因此,科学地选取输入变量方法与评价模型相适配,是提升易发性评价准确性的关键。目前,用于输入变量的方法主要有原始值法^[7]、频率比法^[8](frequency ratio, FR)和邻域频率比法^[9](neighborhood frequency ratio, NFR)。原始值法^[10]直接使用评价指标的原始连续值作为评价模型的输入,具有保真、简便的优势,但其固有的量纲差异和缺乏频率关联的弱点,使其应用效果严重受制于模型的选择,因此该方法并未被广泛应用。FR在离散化连续指标的基础上,统计其子类的崩岗密度(子类崩岗占比与子类面积占比的比值)来进行赋值并作为评价模

型的输入^[11-12],类似的方法还包括证据权法^[13]、信息量法^[14]等均是在离散化连续指标的基础上,建立崩岗与离散化的子类间的数学关系进行赋值的。NFR^[9]是一种新的赋值方法,它通过计算连续指标特定邻域内的崩岗密度进行赋值。郭飞等^[15]选择改进FR和传统FR 2种输入变量输入到随机森林模型(random forest, RF)中进行比较发现,改进FR模型较传统FR模型精度更高;盛明强等^[16]采用FR和原始值2种输入变量发现,FR值输入变量明显优于原始值输入变量;孔嘉旭等^[17]采用离散值和信息量2种输入变量,并分别输入到RF和卷积神经网络中发现,信息量输入变量较离散值输入变量更优;杨文礼等^[18]、白光顺等^[19]采用证据权法进行地质灾害易发性评价取得较好的结果。虽然输入变量方法多样,但不同的机器学习模型适宜的输入变量可能不同,其对崩岗易发性评价结果的影响规律还有待更深入的研究,特别是不同输入变量(如原始值、FR值与NFR值)对机器学习模型性能的影响。

近年来,评价模型方面多选用数据驱动模型,依托遥感和GIS技术,学者应用逻辑回归^[20]、多层感知机(multilayer perceptron, MLP)^[20]、支持向量机^[6]、逻辑回归^[21]、RF^[22]、信息量模型^[23]和人工神经网络^[20]等模型开展崩岗风险评估,取得一系列研究成果。对于崩岗风险评估方法和体系的选择,不同模型各有优点^[24],其中,管家琳等^[25]、杨硕等^[26]、刘永垚等^[27]研究表明,RF模型具有强大的预测能力和稳健性,在处理大量特征、高维度数据时效率极高,尤其针对复杂问题拥有较高的泛化能力,适合开展崩岗易发性评价。此外,对于潜在灾害高风险区识别及大区域尺度,MLP模型的预测精度较高^[28-29]。由于MLP是通过优化网络权重实现建模的典型数值模型,有助于系统揭示原始值、FR

及邻域频率 NFR 等不同输入变量对模型性能的影响机制。同时,作为结构简单且构成深度神经网络基础的人工神经网络,对 MLP 开展针对性研究,能够为后续深化复杂神经网络在灾害易发性评价中的应用提供理论参照与方法基础。不同输入变量对机器学习模型的适宜性不同。原始值法保留了数据的完整信息但存在量纲差异问题;FR 法通过离散化建立了指标与崩岗的数学关系但可能损失连续性信息;NFR 法则在保持连续性的同时建立了概率关联。另外,MLP 模型依赖数值分布进行梯度优化,而 RF 模型基于树结构分裂对数值尺度不敏感,二者对输入变量的适应性存在本质差异。因此,系统比较不同输入变量与机器学习模型的匹配性,揭示其对崩岗易发性评价结果的影响规律,是当前亟需解决的科学问题。

综上,本研究选取江西省石城县作为典型案例,基于 ArcGIS 平台建立了崩岗空间信息数据库,在此基础上,采用地理探测器筛选并构建评价指标体系,并将原始值、FR 和 NFR 作为输入变量,输入到 MLP 和 RF 模型中进行崩岗易发性评价,并开展不同输入变量与机器学习模型对崩岗易发性的适宜性研究。

1 材料与方法

1.1 研究区概况

赣州市是典型的南方红壤区,崩岗发育显著,其崩岗数量与侵蚀面积均居江西省首位。本研究以赣州市东北部的石城县为研究区(25°57'~26°37'N,116°05'~116°39'E),总面积约 1 567 km²,东北部为山地区,西南部为丘陵区,中部则为相对平坦的河谷平原。研究区处于亚热带季风气候区,气候温和,雨水充沛,雨季集中 4—9 月,年平均降水量为 825 mm。区内土壤多

为花岗岩风化形成的红壤土,结构松散,在径流冲刷下,容易造成面蚀和沟蚀,最终在极短的时间内发育成为崩岗。区内共分布崩岗点 2 471 处(图 1),主要集中在研究区西北部的山地丘陵区,与该区域的地形地貌、降雨条件和土壤特性密切相关。

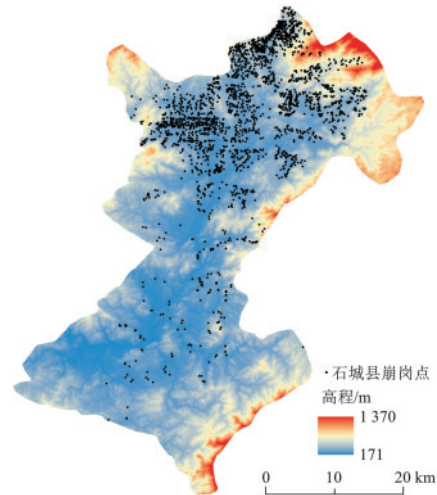


图 1 研究区崩岗点分布

Fig.1 Distribution of Benggang points in study area

1.2 崩岗评价指标选取

崩岗是水力和重力共同作用的土壤侵蚀类型,其发育与地形地貌、土壤气候等方面密切相关,因此本文从多个开源数据库选取 20 个潜在评价指标,涵盖地形地貌、气象水文、土壤性质、植被覆盖等影响崩岗发育的环境条件(图 2),具体包括地形地貌指标(高程、坡度、剖面曲率等)、气象水文指标(降雨侵蚀力、地形湿度指数等)、土壤性质指标(土壤可蚀性、砂粒含量、黏粒含量等)及植被覆盖指标(叶面积指数、森林高度等),具体见表 1。

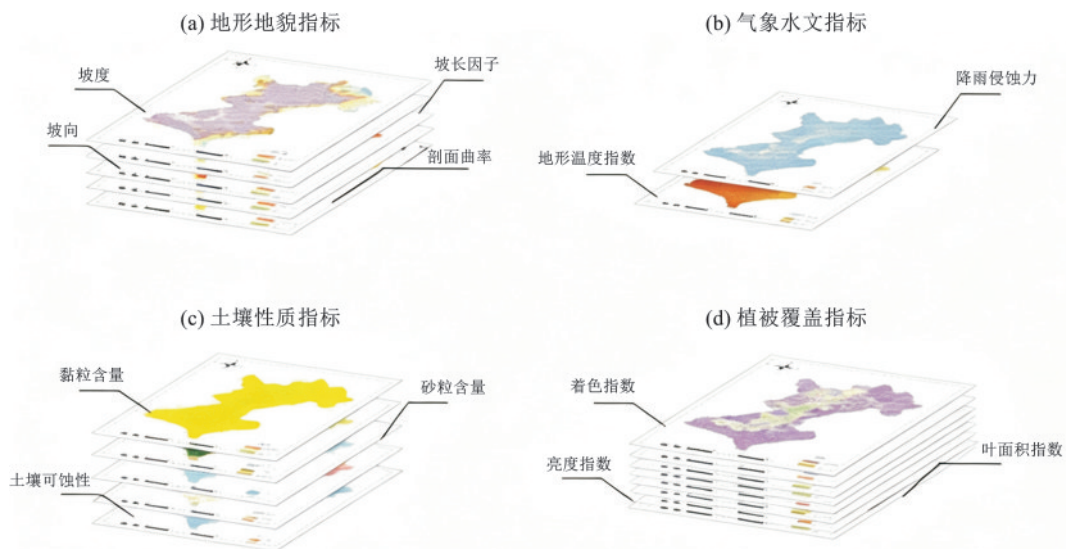


图 2 评价指标

Fig.2 Assessment indicators

表 1 为评价指标的数据来源,包括遥感影像数据、高程数据、土壤数据、降雨数据等。所有数据经过统一的空间配准和重采样处理,栅格

分辨率统一为 30 m×30 m,确保数据的空间一致性。研究区各个评价指标的空间分布特点见图 3。

表 1 研究区评价指标的数据来源

Table 1 Data sources of assessment indicators in study area

数据名称	缩写	数据来源
标准化 VH 通道后向散射系数	VH	来源于 Sentinel-1 卫星,利用 Google Earth Engine 平台处理后提取(https://scihub.copernicus.eu/dhus/#/home)
标准化 VV 通道后向散射系数	VV	
植被覆盖度	FVC	
叶面积指数	LAI	
着色指数	CI	来源于 Sentinel-2 卫星,利用 Google Earth Engine 平台处理后提取(https://scihub.copernicus.eu/dhus/#/home)
亮度指数	BI	
修正土壤调节植被指数	MSAVI	
坡长坡度因子	LS	
坡度/(°)	Slope	
剖面曲率(ProfC)	ProfC	来源于 ALOS-1 卫星,利用 ArcGIS 对 DEM(Digital Elevation Model)处理后提取(https://search.asf.alaska.edu/#/)
坡向/(°)	Aspect	
地形湿度指数	TWI	
高程/m	DEM	
降雨侵蚀力/(MJ·mm·hm ⁻² ·h ⁻¹ ·a ⁻¹)	R	资源学科创新平台(http://www.data.ac.cn/)
土壤可蚀性/(t·hm ² ·h·hm ⁻² ·MJ ⁻¹ ·mm ⁻¹)	K	
黏粒含量/%	Clay	中国科学院资源环境科学与数据中心(https://www.resdc.cn)
砂粒含量/%	Sand	
土壤侵蚀模数/(t·hm ⁻² ·a ⁻¹)	A	
森林高度/m	Forest	https://glad.umd.edu/dataset/gedi
岩性	Lithology	https://geocloud.cgs.gov.cn/
崩岗点		《南方崩岗防治规划(2008—2020年)》

1.3 不同输入变量方法

1.3.1 原始值输入变量 原始值输入变量直接使用评价指标的原始值,对于离散型评价指标,由于其没有数值意义,只有序号意义,因此采用 1, 2, ..., L 的方式对其进行赋值,如岩性,将赋值好的离散型评价指标和连续型评价指标组合就构成原始值输入变量。

1.3.2 频率比值输入变量 FR 模型是在分析研究区崩岗点分布和评价指标类别关系的基础上,借助于地理信息系统平台,结合空间数据对崩岗易发性进行定量评估的分析模型^[30]。其基本思路是通过计算评价指标各子区间对崩岗发育的影响程度来间接实现模型输入变量的确定,表征评价指标各区间对崩岗发生的相对影响程度^[16]。其计算公式^[31]为:

$$FR_i = \frac{N_i/N}{S_i/S} \quad (1)$$

式中: N 和 N_i 分别为研究区总崩岗数(或崩岗面积)和评价指标 X 第 i 个子区间崩岗数(或崩岗面积); S 和 S_i 分别为研究区总面积和评价指标 X 第 i 个子区间面积。

1.3.3 邻域频率比值输入变量 NFR 法最初在 2017 年被首次应用于易发性评价^[32],本研究采用 NFR 来

构建输入变量即作为机器学习模型特征工程的方法是一种新的尝试。相较于 FR 法, NFR 法的优点是:对连续型评价指标使用邻域技术以替代离散化过程,即无需对连续型评价指标进行离散化,从方法上避免离散化过程导致的信息丢失。邻域频率比法的核心思想是先将连续型数据进行归一化,再以归一化后的单指标值为中心统计其固定邻域范围内的崩岗密度^[33],其表达式为:

$$NFR_{j(b)} = \frac{N_{j(b)}/N}{S_{j(b)}/S} \quad (2)$$

$NFR > 1$ 表示该指标值有利于崩岗的发生, $NFR < 1$ 表示该指标值不利于崩岗的发生。NFR 越大,表明该指标值越有利于崩岗的发生。NFR 的核心是确定合适的邻域值,为确定适宜的邻域宽度参数,本研究测试 0.05、0.08、0.10、0.12、0.15 邻域宽度,通过比较不同宽度下模型的 AUC 值和崩岗密度发现,0.10 时效果最优(AUC=0.860),过小(0.05)导致样本不足,过大(0.15)导致过度平滑,均对模型精度产生不利影响。因此,综合模型判别能力与稳定性,最终选用 0.10 作为邻域宽度进行后续分析。

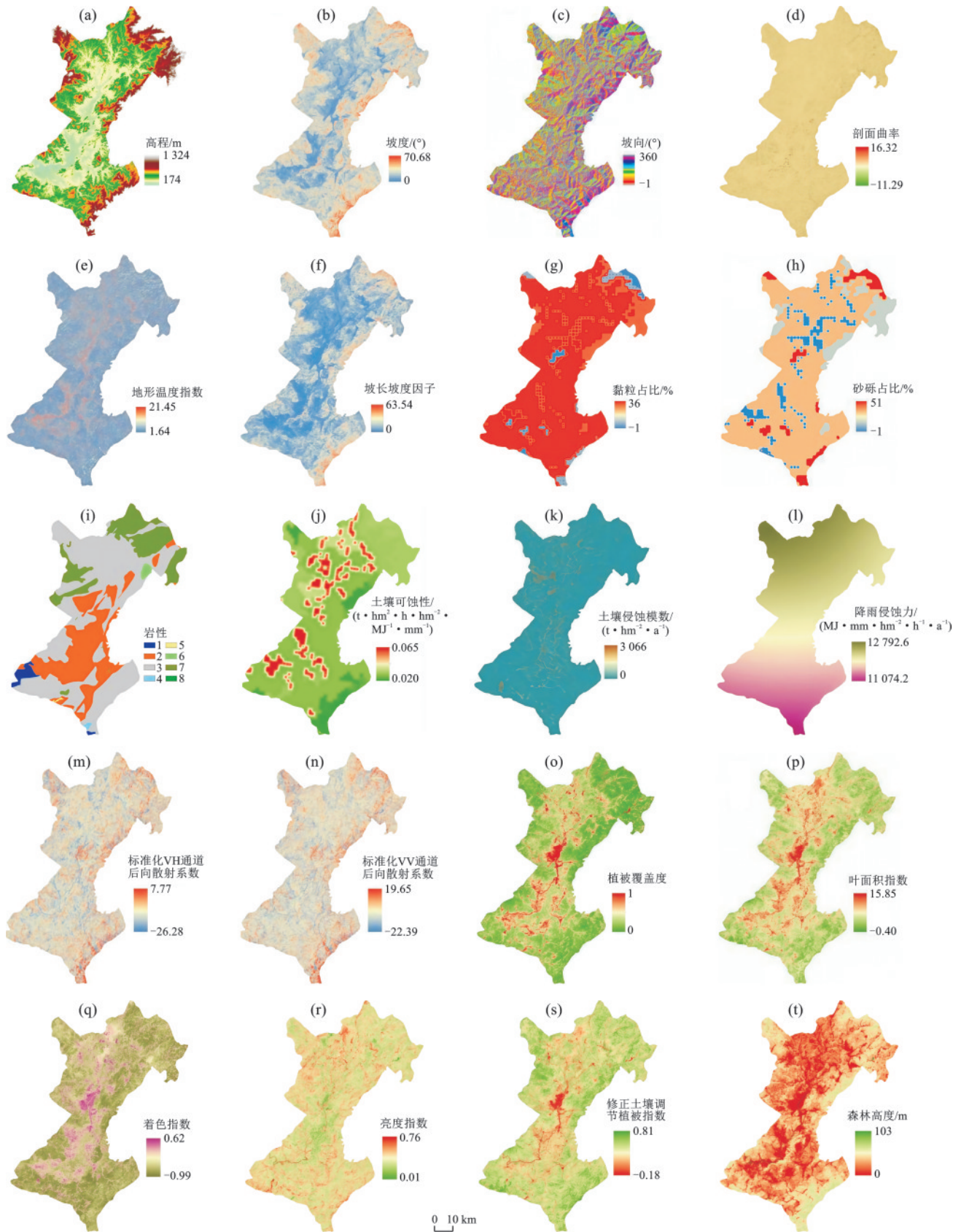


图3 研究区各评价指标
 Fig.3 Assessment indicators in study area

1.4 评价模型

1.4.1 随机森林 RF是Breiman最早提出的利用多决策树的集成学习算法^[34]。该方法的原理基于自助重抽

样技术:通过Bootstrap抽样从原始训练集中有放回地随机抽取k个与初始样本规模一致的子样本集,进而构建k个基分类器并产生k种分类结果;最终通过集成投票

机制,根据各记录在 k 次分类中获得的类别归属进行多数表决,从而确定最优分类结果。模型中有几个重要参数,一个是森林中决策树的数量,另一个是决策树的深度,也就是特征的数量,使用更多的特征可以获得更好的精度,但较大的特征会增加计算的时间^[35]。从原始数据中随机重复采集 N 个样本作为训练集,当 N 较大时,每颗决策树在成长过程中只用到 70% 的训练数据,剩下的约 30% 不会出现在 Bootstrap 样本中,即袋外(OOB)数据。在分类树生成后,利用 OOB 数据得出该树的错误分类率,即 OOB 误差,对森林中所有树的 OOB 误差取平均作为模型的泛化误差^[36]。RF 的泛化误差^[37]:

$$p^* \leq \frac{\bar{\rho}(1-s^2)}{s^2} \quad (3)$$

式中: p^* 为 RF 的泛化误差; $\bar{\rho}$ 为决策树间的相关平均值; s 为决策树的平均强度。

本研究中,RF 模型的关键参数主要包括决策树数量($n_estimators$)选取为 100,最大深度(max_depth)设置为 10,最小分裂样本数($min_samples_split$)为 2,最小叶子节点样本数($min_samples_leaf$)为 1。

1.4.2 多层感知机 MLP 神经网络是一种前向结构人工神经网络,包括输入层、中间隐藏层和输出层 3 层全连接结构^[38]。神经网络主要有 3 个基本要素:权重、偏置和激活函数。多层感知机分类器是 MLP 神经网络中实现分类任务的核心组件。该模型通过构造包含二次惩罚项的损失函数作为优化目标,采用 BFGS 方法对该目标函数进行迭代最小化,以此完成具有单隐藏层的多层感知器网络的参数优化与训练。为提高速度,使用逻辑函数的近似版本作为隐藏层的默认激活函数,但可以指定其他激活函数。在输出层,使用 sigmoid 函数进行分类。MLP 分类器具有高度的并行处理、良好的容错性和非常强的自适应、自学习功能,在地质灾害易发性评价中具有广泛的应用前景^[39]。

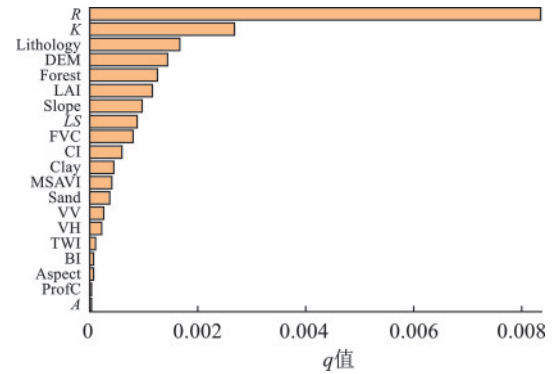
本研究中,MLP 模型采用单隐藏层结构,隐藏层神经元数量设置为 20,激活函数选用 ReLU 函数以增强模型的非线性表达能力,优化器采用 Adam 算法,学习率设置为 0.001,批次大小为 32,最大迭代次数为 500。

2 结果与分析

2.1 评价指标选择

本文采用地理探测器^[40](geodetector, GD)构建指标体系,20 个评价指标中 q 值最大的是降雨侵蚀力(R),最小的是土壤侵蚀模数(A)。为保证崩岗易发性评价的结果,同时降低模型复杂度和计算量,本文依照 q 值从大到小的原则选择累计 q 值占 90% 的指标^[41](图 4),分别为降雨侵蚀力(R)、土壤可蚀性(K)、岩性(Lithology)、高程(DEM)、森林高度

(Forest)、叶面积指数(LAI)、坡度(Slope)、坡长坡度因子(LS)、植被覆盖度(FVC)。9 个评价指标占总指标数的 45%,但其累计 q 值占比达 90%,从而既能保证精度,也可节约计算成本。



注: R 为降雨侵蚀力; K 为土壤可蚀性;Lithology 为岩性;DEM 为高程;Forest 为森林高度;LAI 为叶面积指数;Slope 为坡度; LS 为坡长坡度因子;FVC 为植被覆盖度;CI 为着色指数;Clay 为黏粒占比;MSAVI 为修正土壤调节植被指数;Sand 为砂粒占比;VV 为标准化 VV 通道后向散射系数;VH 为标准化 VH 通道后向散射系数;TWI 为地形湿度指数;BI 为亮度指数;Aspect 为坡向;ProfC 为剖面曲率; A 为土壤侵蚀模数。

图 4 研究区各评价指标的 q 值

Fig.4 q values of assessment indicators in study area

2.2 崩岗易发性评价

由图 5 可知,原始值-MLP 模型将研究区内绝大部分区域划分为极高易发区,表明该模型未能有效区分易发性等级,判别能力较差,即原始值输入变量不适用于 MLP 模型。相比之下,基于 FR 值与 NFR 值输入的 MLP 模型的极高易发区均分布于研究区西北部,说明 2 种输入变量适用于 MLP 模型;然而,NFR 值-MLP 模型的结果在空间上过渡更为自然平滑,而 FR 值-MLP 模型的结果则表现出相对突兀的过渡边界。由图 4 和图 5 可知,基于原始值、FR 值与 NFR 值输入的 RF 模型,其易发性分区整体格局较为一致,仅局部存在细微差异。但 FR 值-RF 模型的结果亦存在边界突变的特征,与 FR 值-MLP 模型类似,但 RF 模型整体上受输入变量影响较小,表明其稳健性优于 MLP 模型。另一方面,原始值输入在 RF 模型中并未导致模型失效,说明不同机器学习模型对输入变量的敏感性存在差异,其影响规律因模型而异。

由表 2 可知,原始值-MLP 模型从极低易发区到极高易发区的崩岗密度并不是单调递增,再次表明其结果不合理。而 FR 值-MLP 和 NFR 值-MLP 从极低易发区到极高易发区的崩岗密度均是单调递增的,并且高和极高易发区的崩岗密度均分别 >1 ,极低、低和中易发区的崩岗密度均分别 <1 ,表明 FR 值-MLP 和 NFR 值-MLP 的分区结果较为合理。

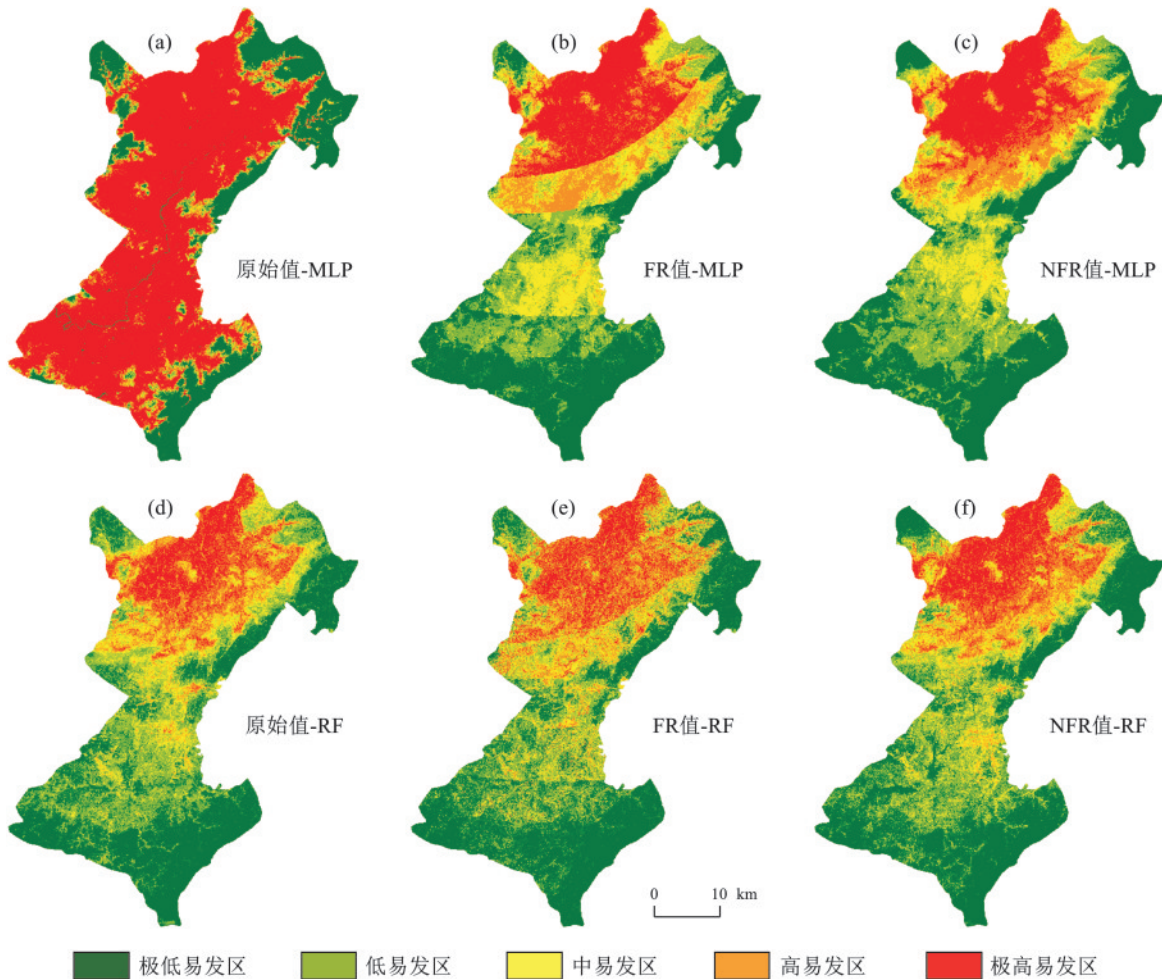


图 5 不同输入变量下的 MLP 和 RF 模型的易发性分区

Fig.5 Susceptibility zoning of MLP and RF models under different input variables

表 2 不同输入变量 MLP 易发区崩岗占比

Table 2 Proportions of Benggang in susceptibility areas of MLP model under different input variables

模型	易发性分区	网格单元数	崩岗数	崩岗密度	极高和高易发区崩岗密度
原始值-MLP	极低	341 161	80	0.17	1.28
	低	63 766	39	0.44	
	中	62 139	55	0.64	
	高	97 943	85	0.63	
	极高	1 135 145	2 093	1.33	
FR 值-MLP	极低	547 863	28	0.04	2.82
	低	359 208	99	0.20	
	中	280 475	222	0.57	
	高	202 446	405	1.45	
	极高	310 162	1 598	3.72	
NFR 值-MLP	极低	570 365	26	0.03	2.96
	低	383 398	110	0.21	
	中	253 312	197	0.56	
	高	210 395	470	1.61	
	极高	282 684	1 549	3.96	

由表 3 可知,原始值-RF、FR 值-RF 和 NFR 值-RF 模型从极低到极高崩岗易发性的崩岗密度均逐渐增大,在极高、高易发区中的崩岗密度均分别 ≥ 1 ,在极低、低、中易发区的崩岗密度均分别 < 1 ,表明它们的分区结果均较为合理。原始值-RF、FR 值-RF 和 NFR 值-RF 模型的极高和高易发区的崩岗密度分别为 3.93、3.69 和 3.83,表明原始值-RF 输入变量的分区结果最好。

从 ROC 曲线可以对结果进行验证。由图 6 可知,原始值、FR 值和 NFR 值的 AUC 值分别为 0.642、0.850 和 0.854,NFR 值-MLP 的 AUC 值略微高于 FR 值-MLP 的 AUC 值。综上所述,NFR 值-MLP 优于 FR 值-MLP 模型,原始值-MLP 模型的结果失效。

3 讨论

本研究通过对比原始值、FR 和 NFR 3 种输入变量形式,并结合 RF 与多层感知机(MLP)模型,对崩岗易发性评价中变量与模型的适宜性进行分析。

表 3 不同输入变量 RF 易发区崩岗占比
Table 3 Proportions of Benggang in susceptibility areas of RF model under different input variables

模型	易发性分区	网格单元数	崩岗数	崩岗密度	极高和高易发区崩岗密度
原始值-RF	极低	594 474	6	0.01	3.93
	低	448 055	45	0.07	
	中	251 895	97	0.28	
	高	211 763	331	1.13	
	极高	193 967	1 873	6.98	
FR值-RF	极低	657 297	17	0.02	3.69
	低	369 211	56	0.11	
	中	257 808	158	0.44	
	高	210 766	378	1.30	
	极高	205 072	1 743	6.14	
NFR值-RF	极低	619 974	15	0.02	3.83
	低	410 333	38	0.07	
	中	251 213	79	0.23	
	高	212 966	295	1.00	
	极高	205 668	1 925	6.77	

输入变量的数据构建方式与模型自身机制共同决定评价结果的稳健性。RF 模型总体上较 MLP 模型更适合开展崩岗易发性评价,该结论与周萍等^[34]、曹文庚等^[29]的研究结果基本一致,验证了 RF 模型在易发性评价中的普适性。

在 MLP 模型中,原始值输入表现较差(AUC=

0.642),分析表明,原始值输入时,由于量纲差异和缺乏频率关联,导致梯度传播不稳定,MLP 依赖数值分布进行梯度优化,需要输入变量与目标变量间具有明确的模型关系。但是,FR 与 NFR 通过概率化转换显式构建指标与崩岗的关系,显著提升模型性能(AUC>0.85)。与之相反,在 RF 模型中,原始值输入反而取得最佳识别效果(高易发区崩岗密度达 3.93),该结论与盛明强等^[16]、孔嘉旭等^[17]的研究结论存在一定的差异,可能与研究区特征、样本规模和模型参数设置有关,需要在更多案例中进一步验证。对比发现,原始值输入 2 种模型的差异化评价结果,根本原因为 MLP 模型和 RF 模型的学习机制不同,RF 模型基于树结构的分裂规则,通过信息增益或基尼系数选择最优分裂点,对变量的数值尺度不敏感,因而能更有效地利用原始信息。

NFR 输入变量在跨模型适应性上表现最优。本研究的 NFR-RF 模型(AUC=0.860)优于郭飞等^[15]的改进 FR-RF 模型(AUC=0.847)和管家琳等^[25]的信息量-RF 模型(AUC=0.832),说明 NFR 方法在保持连续性的同时建立概率关联的优越性;在 MLP 与 RF 模型中,该变量均能取得较高的预测精度与合理的易发性分区结果(MLP 的 AUC 为 0.854,RF 模型中崩岗密度达 3.83)。NFR 通过计算指标值邻域内的崩岗密度,避免传统 FR 方法中离散化处理所带来的信息损失,更好地保持评价指标的连续性,使其能够为不同机器学习模型提供质量更优输入特征。因此,本研究认为,NFR 是一种普适性更强的特征构建方法,适用于多种崩岗易发性评价模型。

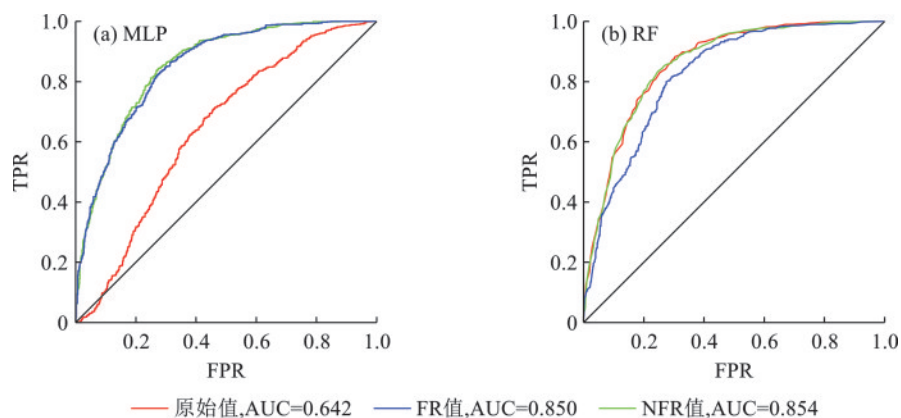


图 6 不同输入变量 MLP 和 RF 模型的 ROC 曲线
Fig.6 ROC curves of MLP and RF models under different input variables

4 结论

1)邻域频率比值输入变量在 MLP 和 RF 模型中的稳健性最高。根据算法原理和数据特点推断,邻

域频率比值输入变量适合所有机器学习模型,是崩岗易发性评价适宜的输入变量。

2)RF 模型总体上较 MLP 模型更适合开展崩岗

易发性评价,3种输入变量下RF模型的崩岗密度最大的为原始值-RF模型,NFR值-RF模型次之,FR值-RF模型最小,分别为3.93、3.83、3.69。

3)采用自然断点法将研究区分为极低易发区、低易发区、中易发区、高易发区和极高易发区。6种工况下分区结果中极高易发区内的崩岗密度最大、各易发区的崩岗数量百分比值由极低易发区向极高易发区逐渐增大,选取的2种模型均能有效评估江西省石城县崩岗易发性。石城县崩岗易发性等级呈由南往北逐渐递增趋势。高易发区、极高易发区主要集中在石城县西北部,与崩岗分布较为吻合。

参考文献:

- [1] 文慧,倪世民,冯舒悦,等.赣南崩岗的发育阶段及部位对土壤水力性质的影响[J].农业工程学报,2019,35(24):136-143.
WEN H, NI S M, FENG S Y, et al. Effects of developmental stages and parts of collapsing gully on soil hydraulic properties in southern Jiangxi[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(24):136-143.
- [2] 廖义善,唐常源,袁再健,等.南方红壤区崩岗侵蚀及其防治研究进展[J].土壤学报,2018,55(6):1297-1312.
LIAO Y S, TANG C Y, YUAN Z J, et al. Research progress on collapsing gully erosion and its control in the red soil region of southern China[J]. Acta Pedologica Sinica, 2018, 55(6):1297-1312.
- [3] 邓羽松,蔡崇法.红壤丘陵区崩岗侵蚀调查与监测防治技术进展[J].土壤学报,2025,62(2):322-333.
DENG Y S, CAI C F. Investigation and monitoring and control technology progress of collapsing gully erosion in red soil hilly region[J]. Acta Pedologica Sinica, 2025, 62(2):322-333.
- [4] 潘峰,文慧,袁芳,等.基于信息量模型和频率比模型的江西省崩岗发育风险对比评估[J].农业工程学报,2025,41(5):107-115.
PAN F, WEN H, YUAN F, et al. Comparative assessment of Benggang development risk in Jiangxi Province based on information value model and frequency ratio model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(5):107-115.
- [5] 夏浪,夏王晓,郭飞,等.主客观赋权法下不同空间分辨率栅格单元对崩岗易发性的影响研究[J].水土保持研究,2025,32(6):208-215.
XIA L, XIA W X, GUO F, et al. Study on the influence of raster units with different spatial resolutions on the susceptibility of collapsing gully under subjective and objective weighting methods[J]. Research of Soil and Water Conservation, 2025, 32(6):208-215.
- [6] 张兴存,蒋玉琳,王玉杰,等.基于多目标优化方法的滑坡易发性评价[J].水土保持学报,2024,38(1):104-112.
ZHANG X C, JIANG Y L, WANG Y J, et al. Landslide susceptibility assessment based on multi-objective optimization method[J]. Journal of Soil and Water Conservation, 2024, 38(1):104-112.
- [7] ADNAN M S G, RAHMAN M S, AHMED N, et al. Improving spatial agreement in machine learning-based landslide susceptibility mapping[J]. Remote Sensing, 2020, 12(20):e3347.
- [8] ARABAMERI A, SAHA S, ROY J, et al. Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash River Watershed, Iran[J]. Remote Sensing, 2020, 12(3):e475.
- [9] 郭飞,赖鹏,黄发明,等.基于知识图谱的滑坡易发性评价文献综述及研究进展[J].地球科学,2024,49(5):1584-1606.
GUO F, LAI P, HUANG F M, et al. Literature review and research progress of landslide susceptibility evaluation based on knowledge graph[J]. Earth Science, 2024, 49(5):1584-1606.
- [10] 赵玉,陈丽霞,付圣.变量取值方式对崩塌易发性的影响[J].地理空间信息,2021,19(12):12-17.
ZHAO Y, CHEN L X, FU S. Influence of variable value methods on collapse susceptibility[J]. Geospatial Information, 2021, 19(12):12-17.
- [11] BUID T, TSANGARATOSP, NGUYENV T, et al. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment[J]. Catena, 2020, 188:e104426.
- [12] PHAMB T, PRAKASHI, BUID T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees[J]. Geomorphology, 2018, 303:256-270.
- [13] 胡杨,张紫昭,林世河.基于证据权与逻辑回归耦合的新疆伊犁河谷地区滑坡易发性评价[J].工程地质学报,2023,31(4):1350-1363.
HU Y, ZHANG Z Z, LIN S H. Landslide susceptibility evaluation in Ili River valley, Xinjiang based on coupling of weights of evidence and logistic regression[J]. Journal of Engineering Geology, 2023, 31(4):1350-1363.
- [14] 陈立华,李立丰,吴福,等.基于GIS与信息量法的北流市地质灾害易发性评价[J].地球与环境,2020,48(4):471-479.

- CHEN L H, LI L F, WU F, et al. Evaluation of geological hazard susceptibility in Beiliu City based on GIS and information value method [J]. *Earth and Environment*, 2020, 48(4): 471-479.
- [15] 郭飞, 赖鹏, 陈洋, 等. 不同环境因子联接方法对崩岗易发性评价的影响[J]. *水土保持通报*, 2022, 42(5): 123-130.
- GUO F, LAI P, CHEN Y, et al. Influence of different environmental factor connection methods on susceptibility evaluation of collapsing gully [J]. *Bulletin of Soil and Water Conservation*, 2022, 42(5): 123-130.
- [16] 盛明强, 刘梓轩, 张晓晴, 等. 基于频率比联接法和支持向量机的滑坡易发性预测[J]. *科学技术与工程*, 2021, 21(25): 10620-10628.
- SHENG M Q, LIU Z X, ZHANG X Q, et al. Landslide susceptibility prediction based on frequency ratio connection method and support vector machine [J]. *Science Technology and Engineering*, 2021, 21(25): 10620-10628.
- [17] 孔嘉旭, 庄建琦, 彭建兵, 等. 基于信息量和卷积神经网络的黄土高原滑坡易发性评价[J]. *地球科学*, 2023, 48(5): 1711-1729.
- KONG J X, ZHUANG J Q, PENG J B, et al. Landslide susceptibility evaluation on the Loess Plateau based on information value and convolutional neural network [J]. *Earth Science*, 2023, 48(5): 1711-1729.
- [18] 杨文礼, 白光顺, 孙滨, 等. 云南哀牢山地区地质灾害因子敏感性和易发性评价: 以新平县为例[J]. *中国地质灾害与防治学报*, 2024, 35(6): 145-152.
- YANG W L, BAI G S, SUN B, et al. Sensitivity of geological hazard factors and susceptibility evaluation in Ailaoshan area, Yunnan Province: A case study of Xinping County [J]. *The Chinese Journal of Geological Hazard and Control*, 2024, 35(6): 145-152.
- [19] 白光顺, 杨雪梅, 朱杰勇, 等. 基于证据权法的昆明五华区地质灾害易发性评价[J]. *中国地质灾害与防治学报*, 2022, 33(5): 128-138.
- BAI G S, YANG X M, ZHU J Y, et al. Evaluation of geological hazard susceptibility in Wuhua District of Kunming based on weights of evidence method [J]. *The Chinese Journal of Geological Hazard and Control*, 2022, 33(5): 128-138.
- [20] 李翠漫, 许贵林, 卢远. 广西桂东南区域崩岗关键影响因子及敏感性研究[J]. *长江科学院院报*, 2020, 37(3): 131-136.
- LI C M, XU G L, LU Y. Study on key influencing factors and sensitivity of collapsing gully in southeast Guangxi [J]. *Journal of Yangtze River Scientific Research Institute*, 2020, 37(3): 131-136.
- [21] WEIY J. Identification of geo-environmental factors on Benggang susceptibility and its spatial modelling using comparative data-driven methods [J]. *Soil and Tillage Research*, 2021, 208: e104857.
- [22] GAYENA, POURGHASEMIH R, SAHAS, et al. Gully erosion susceptibility assessment and management of hazard prone areas in India using different machine learning algorithms [J]. *Science of the Total Environment*, 2019, 668: 124-138.
- [23] 孙昆, 程冬兵, 贺佳杰, 等. 崩岗侵蚀发生风险评估方法的对比研究: 以广东省为例[J]. *中国水土保持*, 2018(3): 51-54.
- SUN K, CHENG D B, HE J J, et al. Comparative study on risk assessment methods of collapsing gully erosion: A case study of Guangdong Province [J]. *Soil and Water Conservation in China*, 2018(3): 51-54.
- [24] LIUZ, WEIY J, CUIT T, et al. Spatial scaling effects of gully erosion in response to driving factors in southern China [J]. *Journal of Geographical Sciences*, 2024, 34(5): 942-962.
- [25] 管家琳, 黄炎和, 林金石, 等. 基于信息量模型与随机森林模型的崩岗风险对比评估[J]. *山地学报*, 2021, 39(4): 539-551.
- GUAN J L, HUANG Y H, LIN J S, et al. Comparative assessment of collapsing gully risk based on information value model and random forest model [J]. *Mountain Research*, 2021, 39(4): 539-551.
- [26] 杨硕, 李德营, 严亮轩, 等. 基于随机森林模型的乌江高陡岸坡滑坡地质灾害易发性评价[J]. *安全与环境工程*, 2021, 28(4): 131-138.
- YANG S, LI D Y, YAN L X, et al. Landslide susceptibility evaluation of high and steep bank slopes in Wujiang River based on random forest model [J]. *Safety and Environmental Engineering*, 2021, 28(4): 131-138.
- [27] 刘永垚, 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 Y Y, DI B F, ZHAN Y, et al. Debris flow susceptibility evaluation based on random forest model: A case study of severe disaster areas in Wenchuan earthquake [J]. *Mountain Research*, 2018, 36(5): 765-773.
- [28] CHEN Y. Spatial prediction and mapping of landslide susceptibility using machine learning models [J]. *Natural Hazards*, 2025, 121(7): 1-19.
- [29] 曹文庚, 潘登, 徐郅杰, 等. 河南省滑坡灾害易发性制图研究: 多种机器学习模型的对比[J]. *地质科技通报*, 2025, 44(1): 101-111.
- CAO W G, PAN D, XU Z J, et al. Research on land-

- slide susceptibility mapping in Henan Province: Comparison of multiple machine learning models [J]. *Geological Science and Technology Bulletin*, 2025, 44(1): 101-111.
- [30] 郭飞, 张晶鑫, 山宏盼, 等. 不同栅格分辨率下赣州市石城县崩岗易发性评价研究[J]. *三峡大学学报(自然科学版)*, 2024, 46(2): 63-70.
- GUO F, ZHANG J X, SHAN H P, et al. Study on susceptibility evaluation of collapsing gully in Shicheng County, Ganzhou City under different grid resolutions [J]. *Journal of China Three Gorges University (Natural Sciences)*, 2024, 46(2): 63-70.
- [31] HONG H, LIU J, BUID T, et al. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China)[J]. *Catena*, 2018, 163: 399-413.
- [32] LI L, LAN H, GUO C, et al. A modified frequency ratio method for landslide susceptibility assessment [J]. *Landslides*, 2017, 14: 727-741.
- [33] ZHANG Y, LAN H, LI L, et al. Optimizing the frequency ratio method for landslide susceptibility assessment: A case study of the Caiyuan basin in the southeast mountainous area of China [J]. *Journal of Mountain Science*, 2020, 17(2): 340-357.
- [34] 周萍, 邓辉, 张文江, 等. 基于信息量模型和机器学习方法的滑坡易发性评价研究: 以四川理县为例[J]. *地理科学*, 2022, 42(9): 1665-1675.
- ZHOU P, DENG H, ZHANG W J, et al. Landslide susceptibility evaluation based on information value model and machine learning methods: A case study of Lixian County, Sichuan [J]. *Scientia Geographica Sinica*, 2022, 42(9): 1665-1675.
- [35] SEVGEN E, KOCAMAN S, NEFESLIOGLU H A, et al. A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and Random Forest [J]. *Sensors*, 2019, 19(18): e3940.
- [36] 吴孝情, 赖成光, 陈晓宏, 等. 基于随机森林权重的滑坡危险性评价: 以东江流域为例[J]. *自然灾害学报*, 2017, 26(5): 119-129.
- WU X Q, LAI C G, CHEN X H, et al. Landslide hazard evaluation based on random forest weights: A case study of Dongjiang River basin [J]. *Journal of Natural Disasters*, 2017, 26(5): 119-129.
- [37] 王世宝, 庄建琦, 樊宏宇, 等. 基于频率比与集成学习的滑坡易发性评价: 以金沙江上游巴塘—德格河段为例[J]. *工程地质学报*, 2022, 30(3): 817-828.
- WANG S B, ZHUANG J Q, FAN H Y, et al. Landslide susceptibility evaluation based on frequency ratio and ensemble learning: A case study of Batang-Dege section in the upper reaches of Jinsha River [J]. *Journal of Engineering Geology*, 2022, 30(3): 817-828.
- [38] 翟文华, 王小东, 吴明堂, 等. 基于频率比模型和随机森林模型耦合的地质灾害易发性评价[J]. *自然灾害学报*, 2023, 32(6): 74-82.
- ZHAI W H, WANG X D, WU M T, et al. Evaluation of geological hazard susceptibility based on coupling of frequency ratio model and random forest model [J]. *Journal of Natural Disasters*, 2023, 32(6): 74-82.
- [39] 王新伟, 张漓黎, 莫德科, 等. 基于信息量和多层感知机分类器模型耦合的平果市斜坡类地质灾害易发性评价[J]. *中国岩溶*, 2023, 42(2): 370-381.
- WANG X W, ZHANG L L, MO D K, et al. Evaluation of susceptibility of slope geological hazards in Pingguo City based on coupling of information value and multi-layer perceptron classifier model [J]. *Carsologica Sinica*, 2023, 42(2): 370-381.
- [40] 王劲峰, 徐成东. 地理探测器: 原理与展望[J]. *地理学报*, 2017, 72(1): 116-134.
- WANG J F, XU C D. Geodetector: Principle and prospect [J]. *Acta Geographica Sinica*, 2017, 72(1): 116-134.
- [41] 郭飞, 蒋广辉, 黄晓虎, 等. 环境因子组合和负样本选取策略对花岗岩区崩岗易发性评价的影响[J]. *农业工程学报*, 2024, 40(1): 191-200.
- GUO F, JIANG G H, HUANG X H, et al. The influence of environmental factor combination and negative sample selection strategy on the susceptibility evaluation of rockfall in granite areas [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(1): 191-200.