

doi: 10.19388/j.zgdzdc.2024.368

引用格式: 杨若君, 夏洋德龙, 夏乐, 等. 基于特征指标筛选与负样本优化的山区滑坡易发性评价[J]. 中国地质调查, 2025, 12(4): 113–123. (Yang R J, Xia Y D L, Xia L, et al. Susceptibility assessment of mountain landslide based on feature indicator screening and negative sample optimization[J]. Geological Survey of China, 2025, 12(4): 113–123.)

基于特征指标筛选与负样本优化的山区 滑坡易发性评价

杨若君^{1,2}, 夏洋德龙^{3*}, 夏乐^{1,2}, 贺秋华^{1,2}, 李明波^{1,2}

(1. 湖南省自然资源事务中心, 湖南长沙 410004; 2. 自然资源部洞庭湖流域生态保护修复工程技术创新中心, 湖南长沙 410004; 3. 湖南科技大学地球科学与空间信息工程学院, 湖南湘潭 411201)

摘要: 滑坡易发性评价是开展灾害监测预警工作的基础, 如何科学、合理地筛选特征指标并优化评价样本仍是当前棘手且易被忽略的问题。以湖南省龙山县为例, 基于斜坡单元提取高程、坡度、坡向等 15 项特征指标, 经主成分分析 (principal component analysis, PCA)、相关性分析与共线性诊断筛选优质指标, 提出一种优化负样本 (optimize negative samples, ONS) 的方法构建评价样本, 然后使用确定性系数-随机森林 (certainty factor-random forest, CF-RF) 模型进行滑坡易发性制图, 并根据受试者工作特征 (receiver operating characteristic, ROC) 曲线与合理性分析检验预测结果准确性。研究表明: ONS-CF-RF 模型能显著提升模型评价精度, 其受试者 ROC 曲线下面积 (area under curve, AUC) 较 CF-RF 模型 AUC 提升了 10.64%; 研究区滑坡高、极高易发区较集中于研究区西北角人类聚集区, 低易发区分布于受人类活动影响较小的高海拔山区。研究成果可为龙山县滑坡灾害防治提供科学指导, 也可为同类型区域滑坡易发性分区提供参考依据。

关键词: 滑坡易发性评价; 特征指标; 负样本; ROC 曲线

中图分类号: P642.22

文献标志码: A

文章编号: 2095-8706(2025)04-0113-11

0 引言

滑坡是湖南省最为常见的地质灾害类型之一, 具有点多面广、易发高发等特点, 常年威胁该省居民的生命和财产安全, 并制约社会经济可持续发展。特别是近年来, 随着山区村民切坡建房、开荒建设等人类活动日益加剧, 滑坡事件频发。据相关资料统计^[1], 自 2018 年以来, 湖南省滑坡数量呈逐年增长趋势。如何精准开展滑坡灾害监测预警, 已然成为政府防灾减灾工作关注的重点。滑坡易发性评价是通过对区域历史滑坡发育环境进行分析, 确定斜坡失稳最佳特征组合, 进而预测区域滑坡发生概率的一种方法。

自 20 世纪 60 年代以来, 随着遥感技术、地理信息系统 (geographic information system, GIS) 以及机器学习等技术迅猛发展, 国内外众多学者开展了一系列区域滑坡灾害评价研究^[2-3], 在评价单元^[4]、评价体系^[5]以及评价模型^[6]等方面取得了优秀成果, 使得滑坡易发性评价技术飞跃式发展, 其评价结果的准确性^[7]与实用性^[8]均得到了显著提升。近年来, 更有学者针对评价指标^[9]与评价样本^[10]展开深入研究, 虽取得了一定的成效, 但受限于评价指标尚未出现统一的选取标准以及评价样本准确性无法准确把控等原因^[11], 评价指标筛选与评价样本优化仍是当前非常棘手且易被忽略的问题^[12-13]。

为此, 本文以湖南省龙山县为研究区, 基于斜

收稿日期: 2024-11-26; 修订日期: 2025-05-16。

基金项目: 湖南省自然资源青年科技项目“湖南省 1:1 万地质灾害调查和风险评价技术标准 (编号: 20240201DZ)”与湖南省自然资源事务中心财政项目“洞庭湖流域生态保护修复工程技术创新中心 2025 年度建设运行”联合资助。

第一作者简介: 杨若君 (1991—), 女, 工程师, 主要从事生态环境与自然资源遥感调查与监测工作。Email: 624861130@qq.com。

通信作者简介: 夏洋德龙 (1988—), 男, 工程师, 主要从事地质灾害评估与风险管理方面的研究。Email: 258441181@qq.com。

坡单元,使用主成分分析(principal component analysis, PCA)、相关性分析与共线性诊断综合筛选滑坡特征指标,提出一种通过简单易发性评价获取优化负样本(optimize negative samples, ONS)的方法,然后使用确定性系数-随机森林(certainty factor-random forest, CF-RF)模型评估研究区滑坡易发程度。研究旨在构建相对科学、准确的区域滑坡易发性评价模型,以期丰富滑坡易发性评价方法,并为龙山县防灾减灾工作提供有效的参考意见和决策思路。

1 研究区概况及数据资料

1.1 研究区概况

龙山县地处云贵高原东南部边缘,坐落于湘鄂渝三省交界处,面积约 3 132.76 km²,是湖南省土家族与苗族聚集较多的地区之一。该县地形错综复杂,地势崎岖不平,总体呈现出自西北向东南逐级降低的趋势,海拔为 175 ~ 1 708 m (图 1),分布多种不同地质时期沉积岩,主要地层包括寒武系、奥陶系、二叠系、三叠系、第四系等,为典型喀斯特地貌。龙山县位于亚热带大陆性湿润季风气候区,境内水资源充沛,是良好的孕灾场所。加之当地人类活动剧烈增长,使得滑坡灾害频发,是湖南省地质灾害应急响应关注的重点区域之一。据湘西土家族苗族自治州 1:50 000 地质灾害详查统计,截至 2017 年底,全县累计发生滑坡事件 163 起,存在潜在滑坡 58 处,其中 95% 的滑坡灾害属于受人类活动影响的小型降雨滑坡,如图 1 所示。

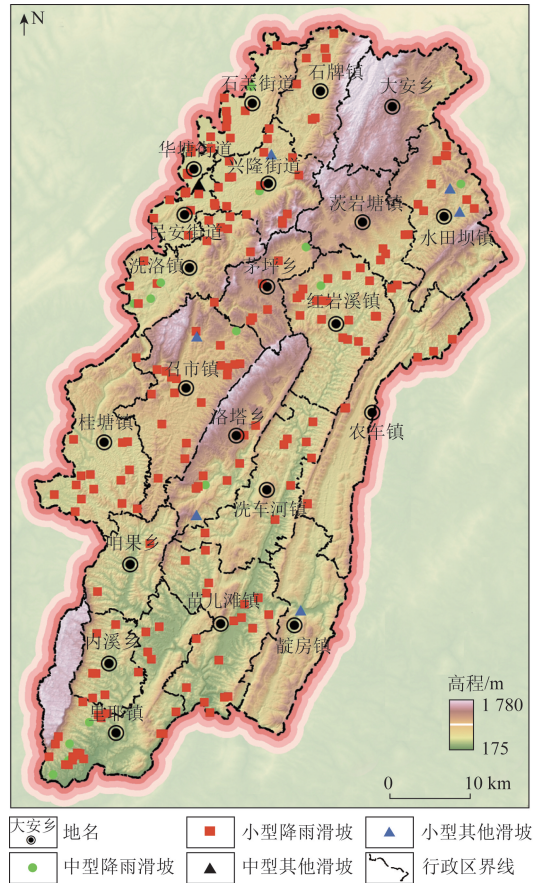


图 1 研究区位置及滑坡分布

Fig. 1 Location of the study area and the distribution of landslides

1.2 数据来源及预处理

本文使用的基础数据包括龙山县历史滑坡灾害点、地质图、行政区划、地质年代与断层数据、河网与路网数据、土地利用数据,以及数字高程模型(digital elevation model, DEM),如表 1 所示。

表 1 资料数据详细来源

Tab. 1 Detailed sources of data

序号	数据名称	来源	类型	备注
1	历史滑坡灾害点	湘西州 1:5 万地质灾害详查	Excel 表格	截至 2017 年,历史滑坡和潜在滑坡共计 221 处
2	地质图	湖南省自然资源事务中心	jpg 图片	纸质扫描存档,比例尺为 1:200 000
3	行政区划	http://datav.aliyun.com	矢量	更新于 2021 年 5 月
4	地质年代与断层	http://dcc.ngac.org.cn	矢量	
5	河网与路网	https://www.resdc.cn	矢量	
6	土地利用	https://gtcd.mnr.gov.cn	栅格	湖南省土地利用图,比例尺为 1:1 200 000
7	DEM	91 卫图	栅格	ALOS DEM,分辨率为 12.5 m

基于 ArcGIS 平台,本文对收集的各项数据资料进行预处理。将收集的龙山县历史滑坡灾害点

与地质图进行矢量化与配准,并统一数据投影坐标(WGS_1984_UTM_49N),构建空间基础数据库。

2 研究方法

2.1 总体研究思路

本文以龙山县为研究区,研究思路如图2所示。首先,在全面收集现有数据基础上,深度挖掘数据信息,以斜坡单元为最小评价单元,提取微观地形指标(高程、坡度、坡向、曲率)、宏观地形指标(地表切割深度、地形粗糙度、高程变异系数)、自然地理指标(森林类型、地形位置指数、太阳光照指数)、水文环境指标(距断层距离、距地表水源距离、地层与地貌)以及人类活动指标(建设用地占比、距

主道路距离)共5类15项特征指标构建初步评价指标数据集。其次,通过PCA、相关性分析与共线性诊断综合筛选优质特征指标,并结合确定性系数(certainty factor, CF)计算指标滑坡分级信息,建立优选评价指标数据集。之后,使用多组随机负样本开展简易滑坡易发性评价,提取公共低易发性区划范围,并在公共低易发区内重新获取非滑坡点作为本研究的ONS数据。最后,结合随机森林(random forest, RF)模型开展滑坡易发性评价制图,采用受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under curve, AUC)与合理性分析检验模型预测精度。

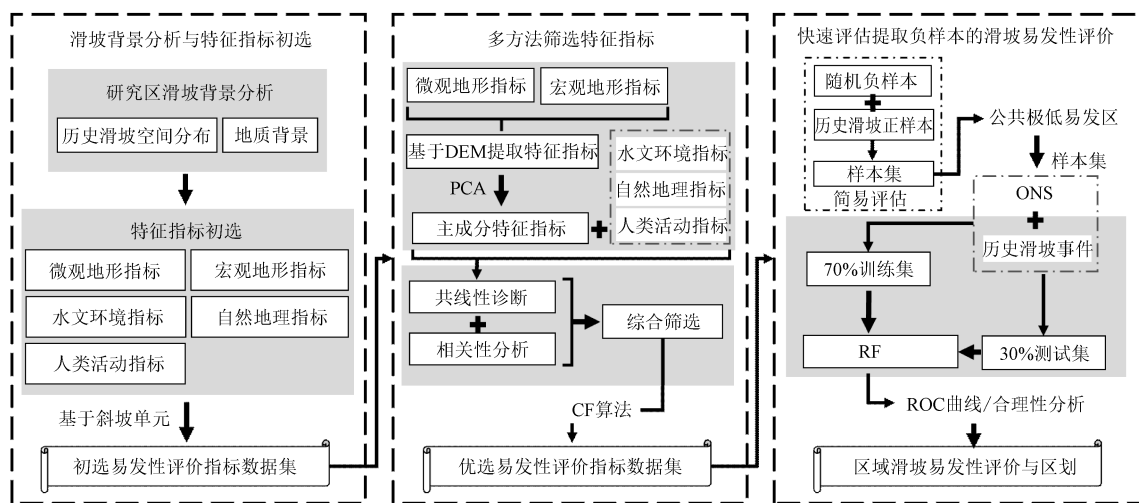


图2 总体技术路线图

Fig.2 Total technology roadmap

2.2 评价单元选择

在滑坡易发性评价中,栅格单元与斜坡单元^[14]是众学者使用频率较高的两类评价单元。其中:栅格单元具备良好的形状规则,能快速剖分与模型计算,从而提升模型计算效率;斜坡单元能依据真实地貌准确的反映区域地形特征,提升滑坡预测评价精度^[15]。为提升滑坡易发性评价精度,本文采用斜坡单元作为最小评价单元。基于GRASS GIS平台,使用r.slopeunits插件对研究区DEM数据划分斜坡单元。在确保每个历史滑坡灾害点有且只有唯一一对对应斜坡单元的前提下,经反复调试,最终以最小流域单元等于1 km²的标准,将研究区划分为63 340个斜坡单元(图3),其中最小斜坡单元面积为0.01 km²,最

大面积为0.56 km²。

2.3 特征指标筛选

2.3.1 PCA法

PCA法是一种致力于数据降维的无监督算法,它通过正交变换将多个线性相关的特征指标转换为少量线性无关的新指标,从而有效去除原指标存在的噪声与冗余信息,提升模型泛化能力^[16]。

考虑到研究区的高程、坡度、坡向、曲率、地表切割深度、地形粗糙度、高程变异系数、地形位置指数以及太阳光照指数9项指标均由龙山县DEM数据计算所得,而森林类型与建设用地占比(来源土地利用数据)、距主道路距离与距地表水源距离(来源河网与路网数据)、地层与地貌和距断层距离(来源地层年代与断层数据)则是通过对应矢量或栅

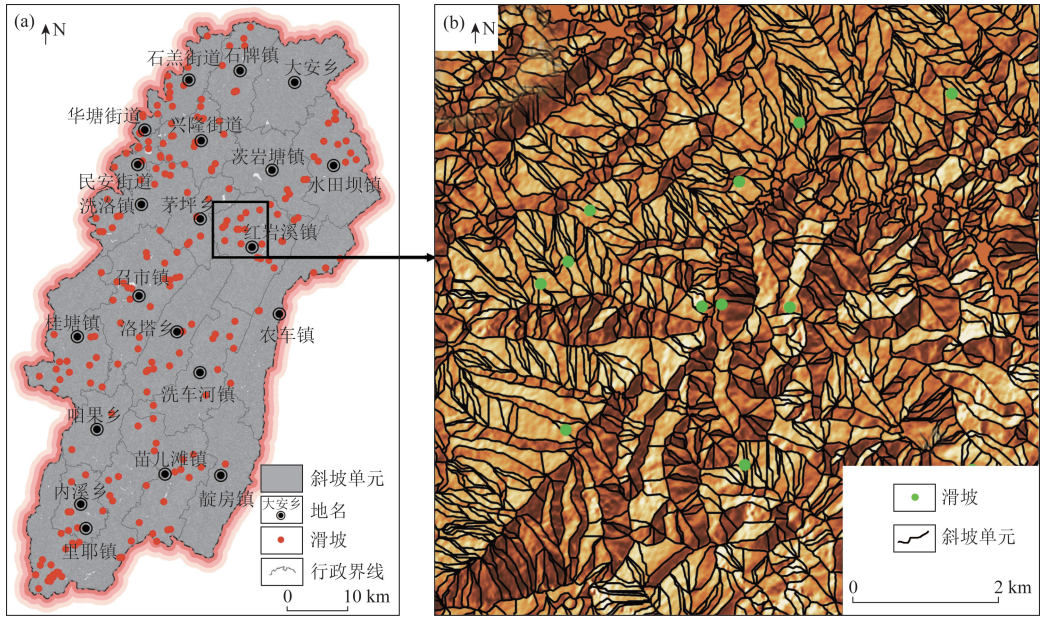


图 3 斜坡单元覆盖区域 (a) 及其细节 (b)

Fig. 3 Coverage area of slope units (a) and their details (b)

格转换所得,故本研究仅对由 DEM 数据获取的 9 项指标开展 PCA。基于 SPSS Pro 软件,将高程、坡度、坡向、曲率、地表切割深度、地形粗糙度、高程变异系数、地形位置指数以及太阳光照指数 9 项指标转换为线性无关的新指标 PC01 至 PC09,并按照方差解释率大于 10%、特征根不低于 0.9 的标准,保留前 4 项特征指标 PC01 至 PC04(表 2)。保留的特征指标总体上既拥有原指标 85.26% 的信息,又缩减了 5 个数据维度。

表 2 成分信息量分析

Tab. 2 Ingredient information analysis

主成分	特征根		
	特征根	方差解释率/%	累积方差解释率/%
PC01	4.216	46.848	46.848
PC02	1.393	15.475	62.323
PC03	1.085	12.055	74.379
PC04	0.979	10.879	85.258
PC05	0.711	7.904	93.161
PC06	0.368	4.091	97.252
PC07	0.182	2.021	99.273
PC08	0.046	0.516	99.789
PC09	0.019	0.211	100

间的关联强度;共线性诊断^[18]可采用方差膨胀因子(variance inflation factor, VIF)检测指标能否被其他指标线性解释。

基于 SPSS 软件,本文将相关性分析与共线性诊断相结合来筛选特征指标,当某指标的 $|R| \leq 0.35$ 且 VIF 值 < 3 时,则认为该指标满足筛选要求。经分析(图 4,表 3),PC02、PC03 与其他指标存在较高关联性 ($|R| > 0.35$),被剔除。最终保留指标为森林类型、距断层距离、距地表水源距离、地层与地貌、建设用地占比、距主道路距离以及 PC01 和 PC04。

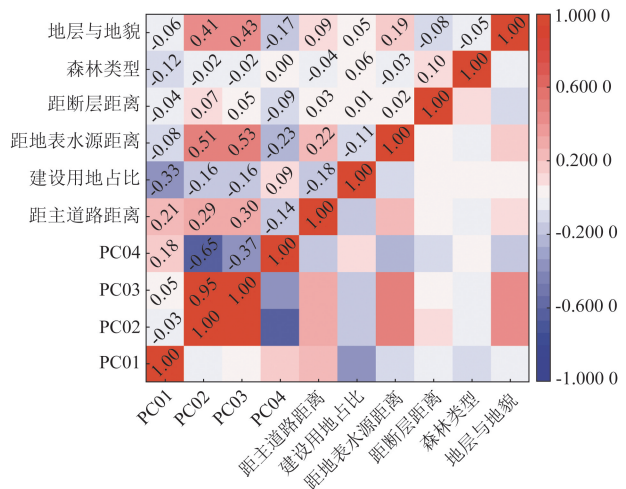


图 4 相关性分析热力图

Fig. 4 Correlation analysis heat map

2.3.2 相关性分析与共线性诊断

相关性分析^[17]可通过相关性系数 R 评估指标

表3 各指标 VIF 值
Tab.3 VIF value for each indicator

	地质时期	森林类型	距主道路距离	建设用地占比	距地表水源距离	距断层距离	主成分 PC01	主成分 PC04
地质时期		1.066	1.068	1.071	1.051	1.068	1.068	1.055
森林类型	1.070		1.075	1.070	1.072	1.073	1.027	1.072
距主道路距离	1.154	1.157		1.158	1.109	1.158	1.084	1.138
建设用地占比	1.005	1.000	1.005		1.005	1.005	1.005	1.005
距地表水源距离	1.133	1.151	1.106	1.155		1.155	1.151	1.109
距断层距离	1.014	1.016	1.018	1.018	1.018		1.014	1.011
PC01	1.163	1.114	1.091	1.166	1.162	1.162		1.129
PC04	1.134	1.148	1.132	1.152	1.107	1.145	1.115	

2.4 评价模型

2.4.1 CF 算法

CF 算法^[19]的本质是一种基于既定事实的条件概率函数。该算法假定历史滑坡发生条件与未来滑坡发生条件相同,来定量计算滑坡灾害特征指标的敏感性,其计算公式为

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_s(1 - PP_a)} & (PP_a < PP_s) \\ \frac{PP_a - PP_s}{PP_a(1 - PP_s)} & (PP_a \geq PP_s) \end{cases} \quad (1)$$

式中: CF 为确定性系数值,表示历史滑坡特征指标的敏感程度,取值范围在 $[-1, 1]$ 之间,当 CF 值越趋近 1 时,代表滑坡发生的敏感性越高,当 CF 值越趋近 -1 时,则代表滑坡发生的敏感性越低,当 CF 值靠近 0 时,表示不能确定滑坡发生的敏感程度; PP_a 为历史滑坡在特征指标的 a 分类中发生的条件概率,常用 a 分类中滑坡数量除以 a 分类面积求得; PP_s 为历史滑坡在整个研究区内发生的先验概率,常用研究区总滑坡数量除以研究区总面积求得。

2.4.2 CF - RF 模型

RF 模型^[20]是一种基于 Bagging (Bootstrap Aggregating) 策略的集成模型,能够有效地处理非线性问题,计算指标权重,被广泛应用于滑坡易发性评价。CF - RF 模型则是将 CF 算法与 RF 模型相耦合得到的一种加权模型,该模型既具备 CF 算法的直观性,又具备 RF 模型的鲁棒性^[21],具体计算公式为

$$LSI = \left(\sum_{i=1}^k \sum_{j=1}^n CF_{ij} \times RF_i \right)_{归一化} \quad (2)$$

式中: LSI 为滑坡易发性指数; CF_{ij} 为第 i 类特征指标中第 j 分级的 CF 值; RF_i 为第 i 类特征指标的权重; k 为特征指标的数量; n 为第 i 类特征指标的分级数量。

通过将式(1)求得的特征指标的分级信息量 CF_{ij} 与基于滑坡样本使用 RF 模型获取的指标权重 RF_i 对应相乘,即可求得特征指标的耦合信息量,然后将指标耦合信息量进行空间叠加与归一化处理,从而得到研究区滑坡易发性指数 LSI 。

2.5 ONS 方法

在滑坡易发性评价中,大多数研究者常采用随机生成非滑坡点的方式获取负样本数据^[22]。该方法虽然具有方便、快捷的特点,但所获取的负样本往往存在噪声,使得相同模型下不同随机负样本得到的评估结果出现明显差异,从而影响评价精度。为此,本文提出一种通过多次简易评价获取 ONS 方法。该方法的具体步骤如下: ①基于对研究区历史滑坡数据的空间分析,选择合适的缓冲区半径建立灾害缓冲区; ②在缓冲区外随机抽取 K 组非滑坡点 ($K=2, 3, 4, \dots$); ③将非滑坡点分别与历史滑坡点组成样本集 K , 并使用统计算法制作易发性评价图; ④空间叠加各组易发性评价图,得到公共低易发区范围; ⑤基于公共低易发区,重新提取非滑坡点以构建优化负样本集。具体流程如图 5 所示。

本文根据研究区滑坡的空间分布状况、规模大小及其影响范围等信息,结合专家经验,最终确定缓冲区半径为 1 km,并随机抽取了 3 组非滑坡点分别通过信息量法、频率比法以及逻辑回归模型制作易发性评价图,得到公共低易发区。

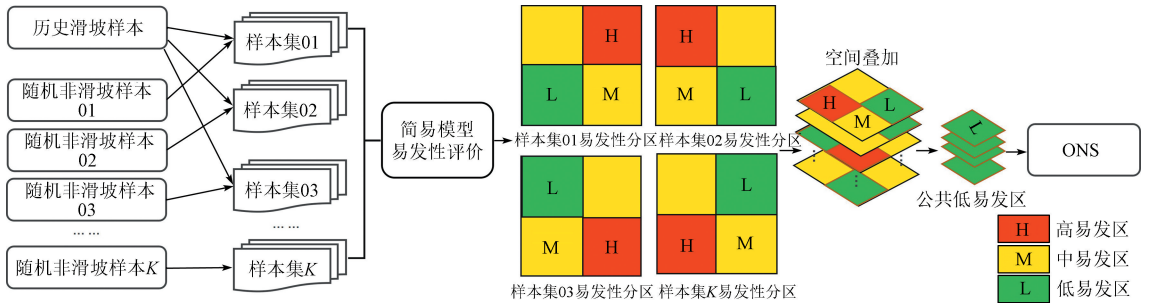


图 5 ONS 方法获取流程

Fig. 5 Acquisition process of ONS method

3 滑坡易发性评价

3.1 特征指标分级

根据研究区滑坡分布规律与指标属性,本研究将特征指标划分为离散型指标(地层与地貌、森林

类型)和连续型指标(距断层距离、距地表水源距离、建设用地占比、距主道路距离以及 PC01、PC04)两类。对于离散型指标,根据其原本属性类别进行分级;对于连续型指标,根据历史滑坡累积变化速率进行分级。具体分级信息(表 4)与分级栅格图(图 6)如下所示。

表 4 研究区特征指标分级信息

Tab. 4 Classification information of feature indicator in the study area

类型	评价指标	分区标准	CF 值	类型	评价指标	分区标准	CF 值
离散型	森林类型	非林地	-0.567 1	连续型	距断层距离	< 300 m	0.400 4
		竹林地	0.408 3			[300,600) m	0.159 2
		灌木林地	-0.038 2			[600,900) m	-0.017 4
		乔木林地	0.013 7			[900,1 200) m	-0.475 1
		其他林地	-0.105 8			[1 200,1 500) m	-0.551 6
	地层与地貌	奥陶系	0.050 2			[1 500,1 800) m	-0.136 1
		志留系	0.378 7			[1 800,2 100) m	-0.440 2
		泥盆系	-1.000 0			≥ 2100 m	0.075 8
		二叠系	-0.792 5			< -26	-0.087 5
		三叠系	-0.767 6			[-26, -24)	-0.120 1
连续型	距主道路距离	第四纪	0.859 5	[-24, -22)	0.374 6		
		白垩系	0.664 8	[-22, -20)	0.126 5		
		寒武系	-0.532 8	[-20, -18)	0.253 8		
		河流积水区	0.662 9	[-18, -16)	0.195 9		
		< 250 m	0.525 2	[-16, -14)	0.167 7		
	距地表水源距离	[250,750) m	-0.070 5	[-14, -12)	-0.140 0		
		[750,1 250) m	0.103 1	[-12, -10)	-0.398 7		
		[1 250,1 750) m	-0.292 3	[-10, -8)	-0.610 1		
		[1 750,2 250) m	-0.331 2	[-8, -6)	-0.749 5		
		[2 250,2 750) m	-0.682 4	≥ -6	-0.716 0		
连续型	建设用地占比	≥ 2 750 m	-0.311 6	PC01	< -40	-0.655 7	
		< 2%	-0.754 5		[-40, -20)	-0.511 5	
		[2,12) %	0.689 8		[-20,0)	0.008 7	
		[12,22) %	0.851 2		[0,20)	0.123 6	
		[22,32) %	0.874 7		[20,40)	0.204 2	
	距地表水源距离	[32,42) %	0.968 2		[40,60)	-0.425 6	
		[42,52) %	0.743 3		[60,80)	0.278 9	
		≥ 52%	0.489 5		[80,100)	0.005 5	
		< 200 m	0.527 7		[100,120)	0.214 1	
		[200,450) m	-0.015 2		[120,140)	-0.167 2	
连续型	距地表水源距离	[450,700) m	0.089 3	[140,160)	0.275 7		
		[700,950) m	-0.086 2	[160,180)	-0.213 2		
		[950,1 200) m	-0.313 3	[180,200)	0.585 0		
		[1 200,1 450) m	-0.601 8	≥ 200	0.078 6		
		[1 450,1 700) m	-0.815 2				
		≥ 1 700 m	-0.877 3				

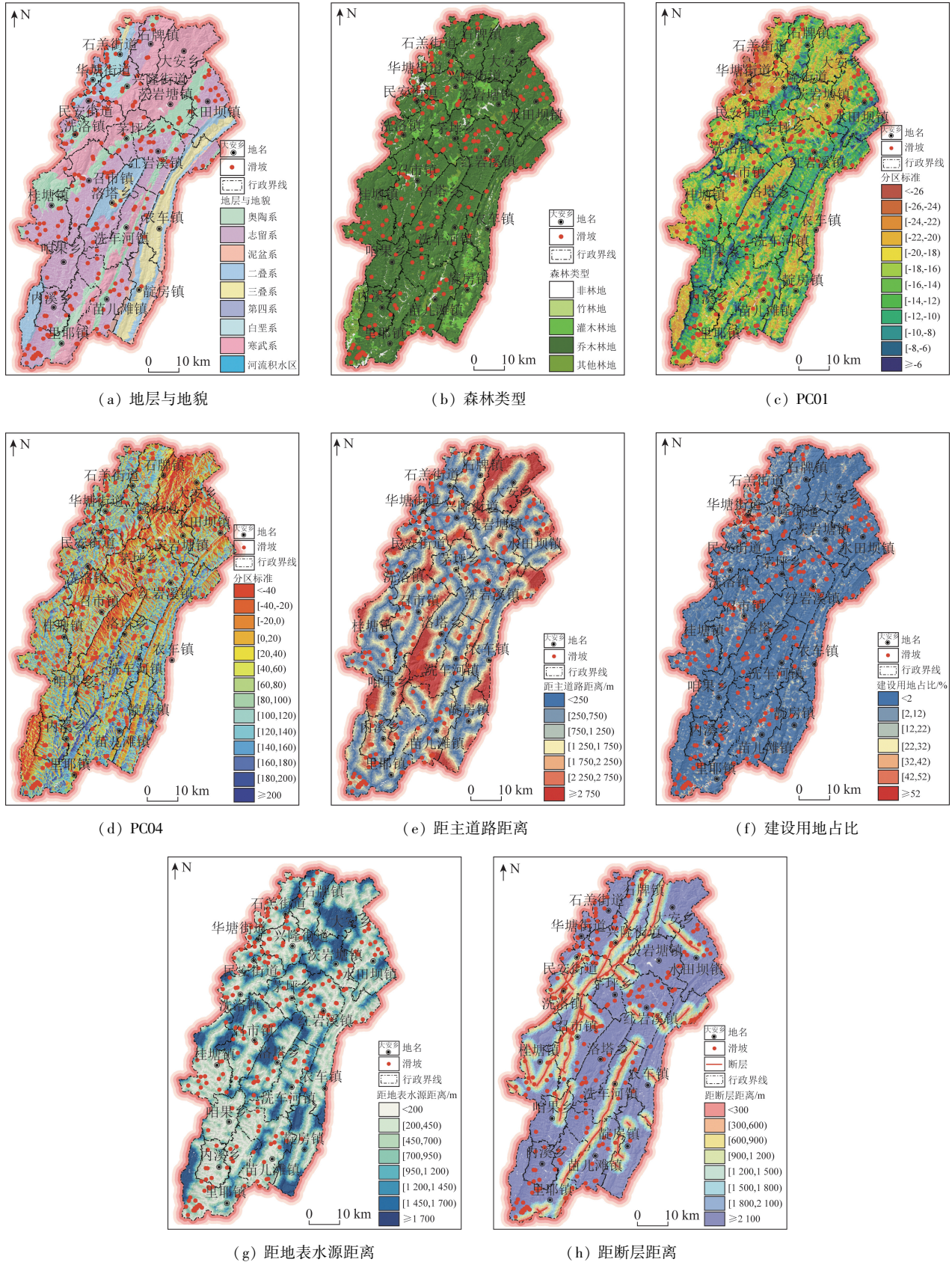


图6 特征指标分级栅格图

Fig. 6 Hierarchical raster plot of the feature indicators

3.2 滑坡易发性制图

将研究区历史滑坡灾害与优化前、后的非滑坡灾害点分别组成样本集,并按照 7:3 比例划分为易发性预测模型的训练集与测试集,输入 RF 模型进行预测,得到各特征指标权重(图 7)。

根据特征指标权重值与 CF 信息量,在 ArcGIS 软件中分别求得 ONS - CF - RF 模型与 CF - RF 模型滑坡易发性指数,并按照相关地质调查技术标准^[23]制作滑坡易发性评价图。通过自然断点法将研究区划分为低易发区、中易发区、高易发区以及极高易发区 4 个等级,结果如图 8 所示。

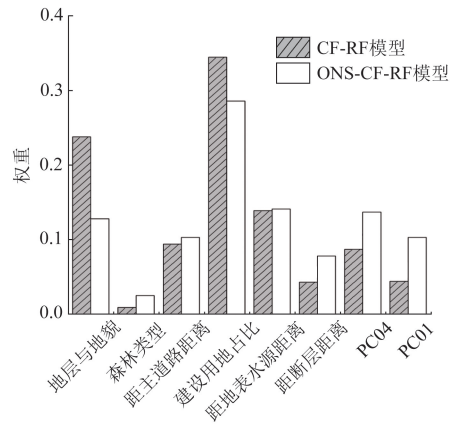
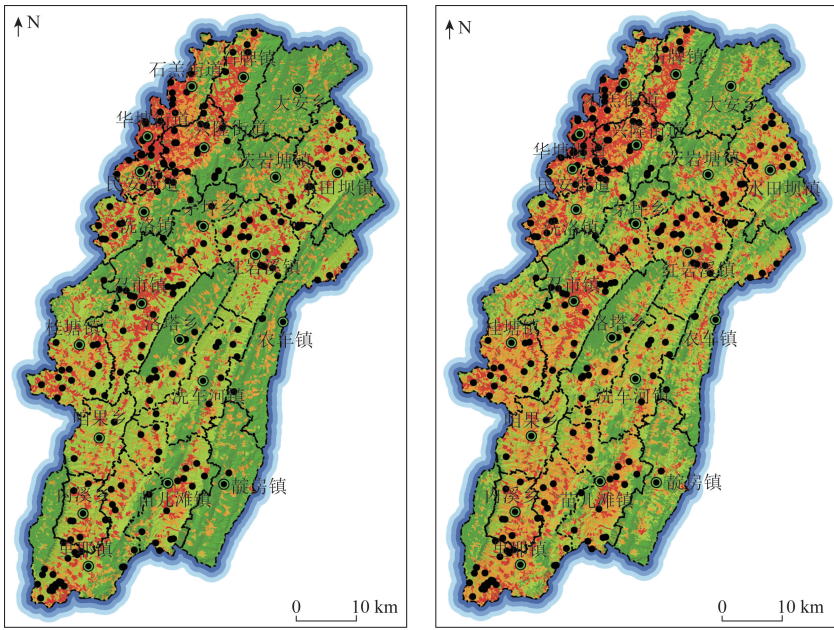


图 7 特征指标权重

Fig. 7 Weights of the feature indicators



(a) ONS - CF - RF 评估结果

(b) CF - RF 评估结果

地名
 滑坡
 行政界线
 低易发区
 中易发区
 高易发区
 极高易发区

图 8 研究区滑坡易发性分区

Fig. 8 Landslide - prone zoning in the study area

3.3 结果检验与分析

3.3.1 合理性分析

将历史滑坡灾害数据分别与两组模型滑坡易发性区划进行空间联接并使用频率比法 (frequency

ratio method, FR) 依次计算各模型滑坡易发性区划信息量,得到研究区滑坡分布结果(表 5)。经统计分析,ONS - CF - RF 模型与 CF - RF 模型中各等级区划分布总体保持一致,滑坡数量趋近于指数函

表 5 研究区模型结果对比分析

Tab. 5 Comparative analysis of the model results in the study area

分区	CF - RF 模型			
	区域面积/km ²	滑坡数量/个	滑坡面积/10 ⁴ m ²	FR 值
低易发区	751.348 3	2	1.664 0	0.037 8
中易发区	1 000.134 1	17	38.123 0	0.241 5
高易发区	837.074 2	43	22.800 5	0.730 0
极高易发区	495.120 9	155	143.907 7	4.448 7

续表

分区	ONS - CF - RF 模型			
	区域面积/km ²	滑坡数量/个	滑坡面积/10 ⁴ m ²	FR 值
低易发区	1 086.163 4	8	8.619 0	0.104 7
中易发区	1 060.840 2	23	37.142 0	0.308 1
高易发区	491.194 4	37	21.626 5	1.070 4
极高易发区	445.479 5	149	139.107 7	4.753 0

数分布且累计超过 85% 的滑坡均坐落在高、极高易发区,较符合真实滑坡分布规律,模型可满足合理性要求。ONS - CF - RF 模型的高、极高易发区 FR 值均高于 1^[24],且大于 CF - RF 模型高、极高易发区 FR 值,其评价结果更为科学、准确。

3.3.2 ROC 曲线

ROC 曲线是反映和揭示敏感性与特异性连续变量相互关系的综合指标,被广泛应用于地质灾害易发性评估检验中^[25-27]。当评估模型所对应的 AUC 大于 0.5 时,表明该模型具备预测功能,反之,则不具备预测功能。通常 AUC 越大,模型的预测性能越高。本文通过将两组模型预测结果分别导入 SPSS 软件生成 ROC 曲线来进一步检验模型预测能力,发现 ONS - CF - RF 模型 AUC(0.988) 较 CF - RF 模型 AUC(0.893) 提升了 10.64% (图 9)。结果表明,两组模型均具备良好的预测功能,且 ONS - CF - RF 模型预测性能更高。

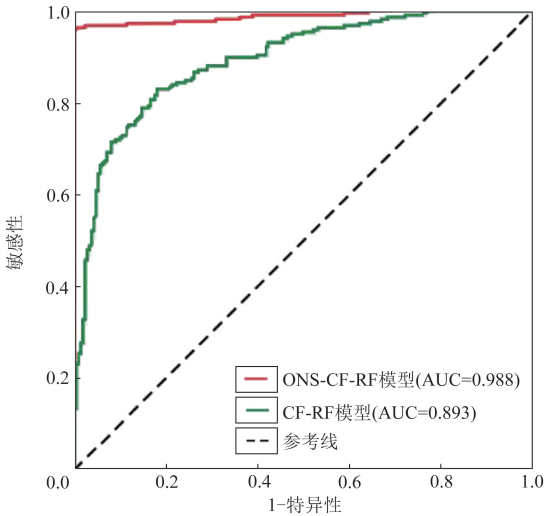


图 9 ROC 曲线

Fig.9 ROC curve

3.3.3 评价结果分析

选取精度更高的 ONS - CF - RF 模型评价结果与研究区土地利用数据进行空间叠加。在剔除真实建设用地范围后,重新统计研究区滑坡易发性分区结果。由表 6 可知,龙山县滑坡灾害低、中、高以

表 6 ONS - CF - RF 模型统计分析

Tab.6 Statistical analysis of ONS - CF - RF model

分区	分区面积 /km ²	分区面积 占比/%	滑坡数 量/个	滑坡密度/(个· 10 000 m ⁻²)
低易发区	1 168.1	37.88	8	0.68
中易发区	1 057.3	34.29	23	2.18
高易发区	467.2	15.15	37	7.92
极高易发区	391.0	12.68	149	38.10

及极高易发区面积分别占研究区总面积的 37.88%、34.29%、15.15% 和 12.68%。整个研究区以中、低易发区为主,极高易发区滑坡密度远超其他分区。高、极高易发区相对集中于龙山县的西北角,该区域人类活动较频繁,地表水资源丰富,地层以白垩系为主,为滑坡发育提供了良好的环境,宜重点关注区域内斜坡发育趋势。人类活动相对匮乏的高海拔山区,地层多为二叠系、寒武系沉积岩,区域内滑坡较难发育,为低易发区。

4 结论

(1) 本文基于 ONS 方案,采用耦合模型对龙山县滑坡易发性进行评估研究。由评估结果可知,ONS - CF - RF 模型 AUC(0.988) 较 CF - RF 模型 AUC(0.893) 提升了 10.64%,拥有更佳的滑坡预测性能,且 ONS - CF - RF 模型的高、极高易发区 FR 值均高于 1,预测结果更科学、合理。研究表明,ONS 方法确实能有效改善模型整体精度,提升研究区滑坡易发性评价的准确性。

(2) 对 ONS - CF - RF 模型滑坡易发性分区结果进行分析,发现研究区滑坡以中、低易发区为主,高、极高易发区主要分布于研究区人类活动更为频繁的西北角白垩纪地层,低易发区主要分布在人类活动较少的高海拔山区。研究成果可为龙山县滑坡灾害防治提供科学指导,也可为同类型区域滑坡易发性分区提供参考依据。

(3) 本文选取湖南省龙山县滑坡易发性进行研究,可为当地灾害防控提供参考,但也存在不足之处,例如研究过程中未对历史滑坡事件的真实性与

准确性进行核实,在未来研究中可着重探讨如何改善滑坡正样本的精度问题。

参考文献(References):

- [1] 国家统计局. 中国统计年鉴 2023 [M]. 北京: 中国统计出版社, 2024: 257.
National Bureau of Statistics of China. China Statistical Yearbook 2023 [M]. Beijing: China Statistics Press, 2024: 257.
- [2] 许强, 郭晨, 董秀军. 地质灾害航空遥感技术应用现状及展望 [J]. 测绘学报, 2022, 51(10): 2020 - 2033.
Xu Q, Guo C, Dong X J. Application status and prospect of aerial remote sensing technology for geohazards [J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(10): 2020 - 2033.
- [3] 李正, 冷亮, 孙永鑫, 等. 基于信息量 - 机器学习耦合模型的水电梯级开发流域滑坡易发性评价 [J]. 测绘通报, 2024(S1): 237 - 241.
Li Z, Leng L, Sun Y X, et al. Landslide susceptibility assessment in the river cascade development basin based on the IV - LM coupling model [J]. Bulletin of Surveying and Mapping, 2024(S1): 237 - 241.
- [4] 安雪莲, 密长林, 孙德亮, 等. 基于不同评价单元的三峡库区滑坡易发性对比——以重庆市云阳县为例 [J]. 吉林大学学报: 地球科学版, 2024, 54(5): 1629 - 1644.
An X L, Mi C L, Sun D L, et al. Comparison of landslide susceptibility in three gorges reservoir area based on different evaluation units: Take Yunyang County in Chongqing as an example [J]. Journal of Jilin University (Earth Science Edition), 2024, 54(5): 1629 - 1644.
- [5] 唐亚明, 薛强, 毕俊攀, 等. 陕北黄土崩塌灾害风险评价指标体系构建 [J]. 地质通报, 2012, 31(6): 979 - 988.
Tang Y M, Xue Q, Bi J B, et al. The construction of factors for assessing the risk of collapse at loess slopes in northern Shaanxi Province [J]. Geological Bulletin of China, 2012, 31(6): 979 - 988.
- [6] Jaafari A. Landslide susceptibility assessment using novel hybridized methods based on the support vector regression [J]. Ecological Engineering, 2024, 208: 107372.
- [7] Ciurleo M, Cascini L, Calvello M. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils [J]. Engineering Geology, 2017, 223: 71 - 81.
- [8] Hussin H Y, Zumpano V, Reichenbach P, et al. Different landslide sampling strategies in a grid - based bi - variate statistical susceptibility model [J]. Geomorphology, 2016, 253: 508 - 523.
- [9] Šilhán K. A new tree - ring - based index for the expression of spatial landslide activity and the assessment of landslide hazards [J]. Geomatics, Natural Hazards and Risk, 2021, 12(1): 3409 - 3428.
- [10] 洪浩源, 王德生, 朱阿兴. 面向机器学习型区域滑坡易发性评价的训练样本采样方法 [J]. 地理学报, 2024, 79(7): 1718 - 1736.
Hong H Y, Wang D S, Zhu A X. A new training data sampling method for machine learning - based landslide susceptibility mapping [J]. Acta Geographica Sinica, 2024, 79(7): 1718 - 1736.
- [11] 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 mapping based on knowledge graph [J]. Earth Science, 2024, 49(5): 1584 - 1606.
- [12] Aleotti P, Chowdhury R. Landslide hazard assessment: Summary review and new perspectives [J]. Bulletin of Engineering Geology and the Environment, 1999, 58(1): 21 - 44.
- [13] 罗路广, 裴向军, 崔圣华, 等. 九寨沟地震滑坡易发性评价因子组合选取研究 [J]. 岩石力学与工程学报, 2021, 40(11): 2306 - 2319.
Luo L G, Pei X J, Cui S H, et al. Combined selection of susceptibility assessment factors for Jiuzhaigou earthquake - induced landslides [J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(11): 2306 - 2319.
- [14] Xia D, Tang H M, Glade T, et al. KNN - GCN: a deep learning approach for slope - unit - based landslide susceptibility mapping incorporating spatial correlations [J]. Mathematical Geosciences, 2024, 56(5): 1011 - 1039.
- [15] 李霞, 宿星, 张满银, 等. 基于证据权法与多源数据的陇中生态脆弱区滑坡敏感性评价——以天水市为例 [J]. 冰川冻土, 2023, 45(1): 67 - 79.
Li X, Su X, Zhang M Y, et al. Landslide susceptibility using weights - of - evidence approach and multi - source data in Longzhong ecologically vulnerable area: A case study of Tianshui City [J]. Journal of Glaciology and Geocryology, 2023, 45(1): 67 - 79.
- [16] 姚翔曦, 张英, 张国治, 等. 基于数据扩充和故障特征优化的 SCNGO - SVM - AdaBoost 变压器故障诊断技术 [J]. 南方电网技术, 2025, 19(6): 14 - 25.
Yao X X, Zhang Y, Zhang G Z, et al. SCNGO - SVM - AdaBoost transformer fault diagnosis technology based on data augmentation and fault feature optimization [J]. Southern Power System Technology, 2025, 19(6): 14 - 25.
- [17] 夏辉, 殷坤龙, 梁鑫, 等. 基于 SVM - ANN 模型的滑坡易发性评价——以三峡库区巫山县为例 [J]. 中国地质灾害与防治学报, 2018, 29(5): 13 - 19.
Xia H, Yin K L, Liang X, et al. Landslide susceptibility assessment based on SVM - ANN models: A case study for Wushan County in the Three Gorges Reservoir [J]. The Chinese Journal of Geological Hazard and Control, 2018, 29(5): 13 - 19.
- [18] 田尤, 张佳佳, 殷红, 等. 基于多种数理模型及其融合的察雅县城滑坡易发性评价 [J]. 科学技术与工程, 2024, 24(29): 12452 - 12460.
Tian Y, Zhang J J, Yin H, et al. Landslide susceptibility evaluation in Chaya County based on multiple mathematical models and their fusion [J]. Science Technology and Engineering, 2024, 24(29): 12452 - 12460.
- [19] 王守华, 王睿菘, 孙希延, 等. 基于 CF - CNN - LSTM 模型的滑坡易发性评价 [J]. 自然灾害学报, 2024, 33(5): 84 - 95.
Wang S H, Wang R S, Sun X Y, et al. Landslide susceptibility evaluation based on CF - CNN - LSTM model [J]. Journal of Natural Disasters, 2024, 33(5): 84 - 95.
- [20] 葛睿雅, 李晓晖, 袁峰, 等. 基于大数据和熵权 - 随机森林的城市地下空间需求评价 [J]. 合肥工业大学学报: 自然科学版, 2025, 48(3): 360 - 368.
Ge R Y, Li X H, Yuan F, et al. Urban underground space demand evaluation based on big data and entropy - random forest [J]. Journal of Hefei University of Technology (Natural Science), 2025, 48(3): 360 - 368.

- [21] Yuan X Y, Liu C, Nie R H, et al. A comparative analysis of certainty factor – based machine learning methods for collapse and landslide susceptibility mapping in Wenchuan County, China[J]. *Remote Sensing*, 2022, 14(14):3259.
- [22] 王伟,宋月,黄莉,等. 粤港澳大湾区复合灾害系统脆弱性评估[J/OL]. *西南交通大学学报*, (2024-09-03). <http://kns.cnki.net/kcms/detail/51.1277.u.20240903.0915.002.html>. Wang W, Song Y, Huang L, et al. Vulnerability assessment of composite disaster systems in the Greater Bay Area[J/OL]. *Journal of Southwest Jiaotong University*, (2024-09-03). <http://kns.cnki.net/kcms/detail/51.1277.u.20240903.0915.002.html>.
- [23] 自然资源部中国地质调查局. DD 2019-08 地质灾害调查技术要求(1:50 000)[S]. 北京:中国地质调查局, 2019. China Geological Survey. DD 2019-08 Technical Requirement for Geo-hazard Survey (1:50 000)[S]. Beijing: China Geological Survey, 2019.
- [24] 何哲,石玉玲,李富春,等. 基于 LightGBM 模型的甘肃省临夏县滑坡易发性评价[J]. *水资源与水工程学报*, 2024, 35(1): 197-205, 216. He Z, Shi Y L, Li F C, et al. Landslide susceptibility assessment based on LightGBM model in Linxia County, Gansu Province[J]. *Journal of Water Resources & Water Engineering*, 2024, 35(1): 197-205, 216.
- [25] 夏洋德龙,罗伟奇,高俊华,等. 基于逻辑回归的择优耦合模型滑坡易发性评估[J]. *地理空间信息*, 2025, 23(6):61-64, 78. Xia Y D L, Luo W Q, Gao J H, et al. Preferred coupled model landslide susceptibility evaluation based on logistic regression[J]. *Geospatial Information*, 2025, 23(6):61-64, 78.
- [26] 陈建平,辛亚波,王泽鹏,等. 样本选取对地质灾害易发性评价的影响——以山西柳林县为例[J]. *中国地质灾害与防治学报*, 2024, 35(3):152-162. Chen J P, Xin Y B, Wang Z P, et al. Effect of sample selection on the susceptibility assessment of geological hazards: A case study in Liulin County, Shanxi Province[J]. *The Chinese Journal of Geological Hazard and Control*, 2024, 35(3):152-162.
- [27] 黄永芳,郭永刚,黄艳婷. 基于加权信息量和加权确定性系数的藏东南滑坡易发性评价[J]. *中国地质调查*, 2024, 11(3): 108-116. Huang Y F, Guo Y G, Huang Y T. Landslide susceptibility assessment in the southeastern Tibet based on weighted informativeness and weighted certainty factor[J]. *Geological Survey of China*, 2024, 11(3):108-116.

Susceptibility assessment of mountain landslide based on feature indicator screening and negative sample optimization

YANG Ruojun^{1,2}, XIA Yangdelong³, XIA Le^{1,2}, HE Qihua^{1,2}, LI Mingbo^{1,2}

- (1. *Hunan Center of Natural Resources Affairs, Changsha Hunan 410004, China*; 2. *Technology Innovation Center for Ecological Conservation and Restoration in Dongting Lake Basin, Ministry of Natural Resources, Changsha Hunan 410004, China*, 3. *School of Earth Sciences and Spatial Information Engineering, Hunan University of Science and Technology, Xiangtan Hunan 411201, China*)

Abstract: Landslide susceptibility assessment is the basis of disaster monitoring and early warning. How to scientifically and reasonably screen feature indicators and optimize assessment samples is still a difficult and easily ignored problem. The authors took Longshan County of Hunan Province as an example, and screened high-quality indicators by principal component analysis (PCA), correlation analysis and collinearity diagnosis, based on 15 feature indicators such as elevation, slope and slope direction. An optimize negative samples (ONS) method was proposed to construct assessment samples, and then certainty factor – random forest (CF – RF) model was used to map landslide susceptibility. The accuracy of prediction results was tested according to receiver operating characteristic (ROC) curve and rationality analysis. The results show that ONS – CF – RF model can significantly enhance the accuracy of model assessment. The area under curve (AUC) of this model has increased by 10.64% compared to the AUC of CF – RF model. The high landslide prone area is concentrated in human gathering area in the northwest corner of the study area, while the low landslide prone area is distributed in the high-altitude mountainous area which is less affected by human activities. The research results could provide scientific guidance for the prevention and control of landslide disaster in Longshan County, and could also provide references for the landslide prone zoning in the same type of region.

Keywords: landslide susceptibility assessment; feature indicator; negative sample; receiver operating characteristic curve