



基于优化随机森林模型的藏东南地区滑坡主控因子分析及易发性评价

卫璐宁¹, 郭永刚¹, 周兴波²

1. 西藏农牧学院 水利土木工程学院, 西藏 林芝 860000; 2. 水电水利规划设计总院, 北京 100120

摘 要: 以藏东南地区为研究对象, 通过优化随机森林模型对滑坡易发性进行深入分析, 探讨影响滑坡发生的主要因素。通过现场调查、遥感数据分析和文献综述, 系统筛选了滑坡与非滑坡样本, 并优化了模型的滑坡-非滑坡样本筛选方法、影响因子的选取、联结方法的应用以及超参数的优化。采用非支配排序算法(NSGA-II)优化随机森林模型(多目标优化)与 RF-GA 模型(单目标优化)进行对比分析, 最优准确率、召回率、精确率、F1 四项指标较 RF-GA 模型分别提高了 3.3%、8.7%、3.2%、1.9%。在滑坡易发性分区方面, 高、较高、中易发区面积占比分别提升了 2.7%、3.1%、1.2%。通过绘制 ROC 曲线和计算 AUC 值, 验证了 RF-NSGA-II 模型的高准确性(AUC=0.877)。研究结果显示, 藏东南地区的滑坡易发区主要集中在易贡藏布与帕隆藏布的交汇处以及雅鲁藏布江的大拐弯区域。在影响因子的重要性排序中, 距道路距离、高程和距河流距离排名靠前, 地质构造复杂、断层发育密集以及长期构造活动的影响使得这些区域滑坡发育频繁, 特别是在断层纵横交错、岩石破碎和节理发育明显的高易发区域。

关键词: 滑坡灾害; 优化随机森林模型; 易发性评价; 主控因子; 藏东南

Main controlling factor analysis and susceptibility assessment of landslides in southeastern Tibet based on optimized random forest model

WEI Lu-ning¹, GUO Yong-gang¹, ZHOU Xing-bo²

1. School of Water Conservancy and Civil Engineering, Xizang Agriculture and Animal Husbandry University, Linzhi 860000, Tibet Autonomous Region, China; 2. Hydropower Water Conservancy Planning and Design Institute, Beijing 100120, China

Abstract: Taking the southeastern Tibet Region as the research object, the paper analyzes the landslide susceptibility based on optimized random forest (RF) model, and discusses the main influencing factors of landslides. With field survey, remote sensing data analysis and literature review, the study screens the landslide and non-landslide samples systematically, and optimizes the landslide and non-landslide sample screening methods of models, selection of impact factors, application of connection approach and hyperparameters. The random forest model (multi-objective optimization) is optimized by non-dominated sorting genetic algorithm (NSGA-II) and compared with RF-GA model

收稿日期: 2024-07-19; 修回日期: 2024-08-20. 编辑: 张哲.

基金项目: 国家自然科学基金重点支持项目“高原重大工程地质灾害监测与分析”(U21A20158); 西藏自治区科技重大专项“高深峡谷地质灾害防治及对工程安全影响分析关键技术”(XZ202201ZD0003G03); 西藏自治区科技重点研发计划项目“基于大数据下西藏重大水电工程强震监测关键技术”(XZ202201ZY0034G); 西藏农牧学院研究生创新计划“藏东南滑坡实时监测预警系统设计与实现”(YJS2024-50).

作者简介: 卫璐宁(1998—), 男, 硕士研究生, 从事基于大数据的滑坡监测分析预警研究, 通信地址 西藏自治区林芝市巴宜区育才西路 100 号, E-mail//1547406558@qq.com

通信作者: 郭永刚(1966—), 男, 教授, 从事水利水电工程强震安全监测与高原重大工程地质灾害研究, 通信地址 西藏自治区林芝市巴宜区育才西路 100 号, E-mail//1960373107@qq.com

引用格式: 卫璐宁, 郭永刚, 周兴波. 基于优化随机森林模型的藏东南地区滑坡主控因子分析及易发性评价[J]. 地质与资源, 2025, 34(1): 112-127. Wei L N, Guo Y G, Zhou X B. Main controlling factor analysis and susceptibility assessment of landslides in southeastern Tibet based on optimized random forest model [J]. Geology and Resources, 2025, 34(1): 112-127.

(single objective optimization). The four indexes of optimal accuracy, recall rate, precision rate and F1 are increased by 3.3%, 8.7%, 3.2% and 1.9%, respectively, compared with the RF-GA model. Besides, the high, relatively high and medium susceptible areas increased by 2.7%, 3.1% and 1.2%, respectively, in terms of landslide susceptibility zoning. The high accuracy of RF-NSGA-II model is verified ($AUC=0.877$) by drawing ROC curve and calculating AUC value. The results show that the landslide susceptible areas in southeastern Tibet are mainly concentrated in the intersection of Yigong Zangbo River and Palong Zangbo River and the big bend area of Yarlung Zangbo River. In the importance ranking of landslide impact factors, distance from road, elevation and distance from river take up the top places, as complex geological structures, densely developed faults and influence of long-term tectonic activities cause frequent landslides in such areas, especially in high susceptible areas with criss-crossing faults, broken rocks and developed joints.

Key words: landslide disaster; optimized random forest model; susceptibility assessment; main control factor; southeastern Tibet

0 引言

滑坡灾害具有广泛的分布范围、高频的发生率以及严重的灾害损失,对人类财产安全构成了极大的威胁^[1]. 青藏高原东南部因其复杂多变的气候、独特的地质构造和特殊的地形地貌,受构造抬升、地震活动、冰川、气候变迁及人类活动等多重因素的影响,滑坡现象频发^[2-3]. 该地区曾经发生过如易贡大滑坡^[4-6]、帕隆藏布 102 滑坡群^[7]、白格滑坡等典型滑坡事件^[8-9]. 藏东南地区是国家公路、铁路和水电等重大工程建设重点区域. 深入研究致灾因素和机制,细化灾害发生条件,已成为学者和专家们的重要任务^[10].

数据驱动模型因其鲁棒性和高准确性,在滑坡易发性评价中得到广泛应用,经历了数理统计和机器学习两个阶段^[11-12]. 信息量等统计模型通常依赖大量先验知识,计算复杂度较低,未能充分利用高精度数据,存在一定局限性. 在海量数据和计算能力提升的背景下,机器学习模型凭借其在大数据处理上的优势,以及无需过多先验知识的特点,广泛应用于滑坡易发性评价和预测中,并且其成功率和有效性明显高于传统数理统计模型,应用愈加广泛,主要包括支持向量机^[13-14]、随机森林^[13,15]和深度学习模型^[16]等. 在现有的机器学习模型中,随机森林因其使用广泛且效果更佳^[17],在多项研究中相比支持向量机等其他机器学习模型总是能获得更高精度的预测结果^[18]. 然而,常规的机器学习模型仍存在一些问题和挑战. 例如,非滑坡样本的选择不够优化^[19],降低了模型的“学习”效率;模型的可解释性有待加强^[20];传统的因子离散型

分级方式在模型中应用未能充分发挥机器学习的计算优势,导致部分信息丢失,从而降低模型性能^[21];超参数设置往往保留默认参数或通过多次重复试验主观调整,未能客观反映不同地质环境条件下的合理性^[22],导致模型精度下降. 因此,优化机器学习模型在滑坡易发性评价中的应用研究十分必要.

针对现有机器学习模型在滑坡易发性评价中的问题和难点,本文以藏东南部分地区为例,进行模型优化和评价验证研究. 首先,通过基于滑坡面状矢量数据建立缓冲区,改进非滑坡样本的选择方法,利用皮尔逊相关性分析和随机森林基尼系数进行因子筛选,解决信息冗余问题. 此外,在影响因子的定量表征和归一化处理方面,探索解决传统离散分级导致的信息丢失问题. 为了提升模型性能,基于 RF-NSGA-II 算法对随机森林的超参数进行多目标优化. 将经过系统优化的随机森林模型应用于雅鲁藏布江流域的群发滑坡易发性评价中,并与 RF-GA 模型的评价结果进行对比,以期提高模型在滑坡易发性评价中的精度. 通过优化随机森林模型在滑坡易发性预测中的应用,并综合分析其主要控制因素,可识别出滑坡高风险区域,为该地区滑坡灾害的防治提供有效的理论支持^[23-24].

1 研究区概况

研究区位于西藏自治区东南部,地理坐标范围: $94^{\circ}07'—96^{\circ}30'E$, $28^{\circ}02'—31^{\circ}00'N$, 总面积约为 $76\,500\text{ km}^2$,包括波密县、米林县、墨脱县和巴宜区的部分地区(图 1). 地形特征表现为典型的高原丘陵和

高山河谷地貌，主要河流包括雅鲁藏布江、帕隆藏布和易贡藏布，这些河流在地区内形成陡峭的山势和纵横交错的沟谷。地势起伏剧烈，整体呈现出北高南低的特点，主要山脉为东西走向。海拔高度 559~7 782 m，最高峰是南迦巴瓦峰，位于米林县和墨脱县的交界处。

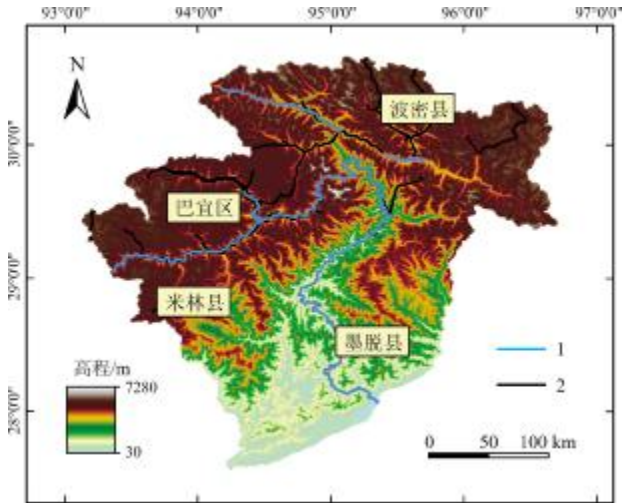


图 1 研究区地理地貌图

Fig. 1 Geomorphologic map of the study area

1—河流 (river); 2—道路 (road); 3—研究区 (study area)

研究区内存在多条主要断裂带，包括巴青—类乌齐断裂、嘉黎—察隅断裂带、鲁朗—易贡断裂及雅鲁藏布江断裂。该区域主要出露的地层包括盆地相上三叠统的砂岩、夹板岩、火山岩，以及海相下一中三叠统的千枚岩、砂岩和含砾状灰岩等。构造变形以挤压逆冲和走滑为主要特征。由于特有的地形特点、高降雨量和板块内部的动力活跃，研究区滑坡灾害发育显著。

2 数据准备

2.1 滑坡编录

滑坡编录数据的完整性与准确性对于滑坡易发性评价至关重要。2018 年 10 月 17—20 日，米林县加拉村段发生山体滑坡，形成堰塞湖，蓄水量约 $3.6 \times 10^8 \text{ m}^3$ ，坝前水深约 59 m。利用 0.75 m 分辨率的灾后 Pleiades 遥感影像，通过人工解译和机器学习分类，对研究区的滑坡灾害进行了遥感编译（图 2），最终获取了 10 452 处滑坡，总滑坡面积达 12.85 km^2 ，占研究区总面积的 7.56%。滑坡规模以小型和中型为主，主要类型包括坡脚侵蚀型土质滑坡和降雨型土质滑坡。

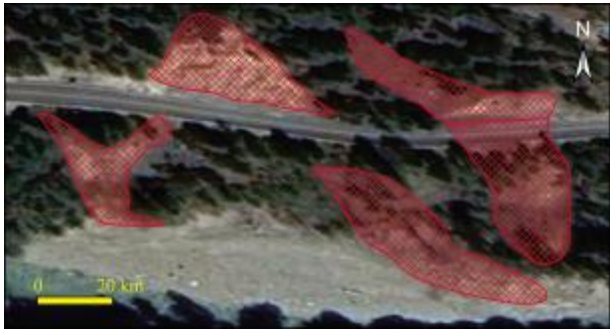


图 2 滑坡遥感影像编译结果

Fig. 2 Interpretation result of remote sensing image

2.2 影响因子

滑坡是由多种因素综合作用而成的地质灾害，其中降雨量、地形地貌和地层岩性等因素是影响其形成的关键因素。不同地区的孕灾条件存在着差异，导致各种因素对滑坡发育的贡献程度不尽相同。地层岩性作为地质灾害发育的物质基础，直接影响着滑坡的形成和演化过程。同时，地形地貌则控制着地质灾害的空间边界条件，为滑坡的发生提供了地理环境基础。河流进一步促进灾害的发生，其位置指示着坡脚侵蚀及坡体的水文地质特征，从而增加了斜坡的不稳定性。地质构造既控制着地形地貌的形成，又影响着岩体结构及组合特征，因此对地质灾害的发育起着综合控制影响作用。根据研究区域的滑坡特征和相关研究结果，初步选择了 17 个影响因子作为滑坡易发性评价的输入变量，其中包括高程、坡度、坡向、地层岩性、断层距离等（见表 1）。这些因子将通过单因素分析进行评估，以揭示它们对滑坡发生的影响程度。

3 随机森林模型优化方法

随机森林(Random Forest, RF)是一种集成学习算法，应用于分类和回归问题。其核心思想是通过集成多棵决策树，提升模型的预测性能和稳定性。首先，随机森林从原始数据集中随机且有放回地抽取多个样本，生成多个训练子集。每个子集用于训练一棵决策树。在决策树的构建过程中，每个节点分裂时仅随机选择一部分特征来决定最佳分裂方式。通过这种样本和特征的随机性，随机森林能够训练出一系列不同的决策树，充分利用数据的多样性。训练完成后，随机森林将所有决策树的结果进行集成。在分类任务中，每棵树独立地对输入样本进行预测，最终通过投票机制

表 1 滑坡灾害影响因子数据来源表

Table 1 Data sources for landslide hazard impact factors

编号	主控因子	评价因子	因子致灾特征及数据来源
1	地形地貌	高程	坡体特征; ASTER GDEM v2.0 (12.5 m 分辨率), 数字地形高程通过 ArcGIS 软件生成
2		坡度	
3		坡向	
4		坡度变率	
5		平面曲率	
6		地形起伏度	
7		坡度变率	
8		坡向变率	
9	地层岩性	地层岩性分区	坡体岩土物理指标; 通过 1:20 万地质图(http://dcc.ngac.org.cn)提取
10	水文	距河流距离	坡体水文因子特征及侵蚀坡脚状况; 通过 1:25 万基础地理(http://dcc.ngac.org.cn)提取
11	人类活动	距道路距离	人类工程设施对坡体影响; 通过 1:25 万基础地理(http://dcc.ngac.org.cn)提取
12	气象	多年有效平均降雨量	降雨入渗对滑坡坡体稳定性影响; 通过国家气候中心获取降雨数据, 1 km 栅格(http://www.ncc-cma.net/cn)
13	地质构造	距主要断层距离	地质构造对坡体发育影响; 基于 1:20 万地质图(公开版)
14	环境地质	土地利用类型	影响坡体发育环境特征; 基于 Landsat 8 遥感影像数据, 30 m 分辨率
15		地表切割深度	
16		地形粗糙度	
17		归一化植被指数	

选择票数最多的类别作为最终预测结果. 在回归任务中, 所有树的预测值取平均值, 得到最终的预测结果. 具体的流程见图 3.

为提升模型的预测精度, 本研究在以下 4 个方面进行了全面优化: 1) 优化样本筛选策略, 确保训练数据集具有更好的代表性和质量, 从而提高模型的基础数据支持; 2) 在影响因子选取方面, 筛选出对模型预测有显著影响的关键因子, 增强模型的有效性和准确性; 3) 优化各个模型组件的联接方法, 改进模型结构, 提高整体性能; 4) 进行关键超参数(树的数量、树的最大深度、最小样本分裂树、最小叶子节点样本数)寻优, 通过不断重复初始种群生成、适应度评估、选择操作、交叉操作、变异操作、迭代更新等步骤, 选择出适应度最高的个体作为最佳超参数组合, 进一步提升模型的预测能力. 为验证优化效果, 将 RF-NSGA-II 和 RF-GA 模型进行了对比分析. 两种模型使用相同的优化方法后输入样本数据集, 通过 ROC 曲线和 4 项常见指

标(准确率、召回率、精确率、F1)对比两种模型的评价结果(图 4). 此优化流程确保了研究区滑坡易发性评价结果的成功率和预测率得到显著提升.

3.1 非滑坡样本筛选策略

本研究采用建立缓冲区的方式, 防止非滑坡样本选取时可能陷入滑坡区域的问题, 以确保所选取的非滑坡样本的可靠性和代表性. 为此, 在 GIS 平台中基于光学遥感影像解译获得的滑坡面状矢量数据, 以滑坡面为基准, 绘制了 50 m 的缓冲区. 保证了所挑选的非滑坡样本位于滑坡边界的外侧, 避免误选. 通过随机抽样的方法从缓冲区中选取等量的非滑坡数据, 从而构建滑坡-非滑坡样本数据集(图 5).

3.2 Gini 指数的度量

在随机森林算法中, Gini 系数被用于属性选择, 以评估各个特征在滑坡发生中的重要性. 通过二分递归分割技术构建简洁的二叉树结构, 从而在选择影响因子时避免了盲目性和主观性. 对于给定的节点因子样

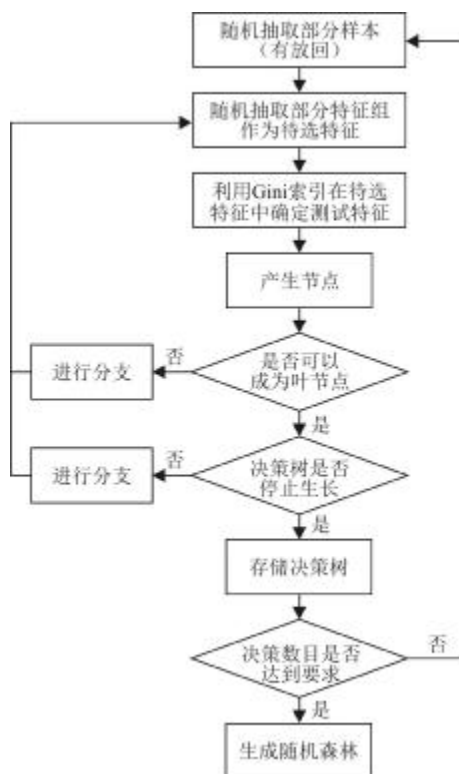


图3 随机森林算法流程图

Fig. 3 Flow chart of random forest algorithm

本集 Q , 计算每个因子 Q 的 Gini 系数, 作为衡量其重要性的指标. Gini 系数反映了数据集的纯度水平, 通过最小化数据集的不纯度, 帮助确定重要因子的特征^[25].

节点中样本 Gini 指数计算公式:

$$\text{Gini}(Q) = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

式中, n 表示因子类别的数量, p_i 表示数据集中属于第 i 类因子的概率.

Gini 系数的取值范围在 0~0.5, 0 表示样本中因子属于同一类, 纯度最高; 0.5 表示样本均匀分布在各个因子中, 不纯度最高. 首先计算当前因子(父节点)的 Gini 不纯度, 对于每个候选分裂点, 计算分裂后的两个因子的加权 Gini 不纯度, 选择能够最大程度降低 Gini 不纯度的节点因子进行分裂. 假设因子节点被分裂成两个子节点 Q_1 和 Q_2 , 则分裂后的加权 Gini 不纯度计算如下:

$$\text{Gini}(Q, T) = \frac{|Q_1|}{|Q|} \text{Gini}(Q_1) + \frac{|Q_2|}{|Q|} \text{Gini}(Q_2) \quad (2)$$

式中, $|Q_1|$ 、 $|Q_2|$ 、 $|Q|$ 分别表示节点 Q_1 、 Q_2 、 Q 中样本数量.

3.3 Pearson 指数计算

Pearson 相关性分析是用于衡量两个服从正态分布变量之间的线性关系的强度和方向. 它通过计算两个变量之间的 Pearson 相关系数来完成, 减少选取影响因素自相关问题.

Pearson 相关系数(ρ), 是衡量两个变量(X, Y)之间线性关系强度和方向的统计量. 通过计算两个因子的协方差与它们各自标准差的商, 取值范围在 -1~+1 之间. 数学上, 它的定义如式(3)所示:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

使用样本的协方差和标准差来估算 Pearson 相关系数, 如式(4), 用 r 表示:

$$r(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (4)$$

3.4 致灾因子量化

致灾因子的量化方法基于微积分概念, 用于避免传统的连续型因子分级离散化. 通过直接提取各连续性因子的实际数值, 快速对数据样本的统计分析, 以此充分体现各因子数据的真实性. 对于离散型因子, 则采用频率比(FR)方式进行定量表征, 以降低其离散性的影响. 在将各影响因子量化后, 进行归一化处理, 以消除不同因子之间的量纲影响, 使得数据处于同一数量级, 从而确保数据处理的方便性, 并使程序运行更加收敛快速.

频率比模型是评估两个变量之间的相关性(概率关系). 其计算公式为:

$$\text{FR}_{ij} = \frac{w_{ij}/W}{s_{ij}/S} - 1 \quad (5)$$

式中, i 为因子类别, j 为因子等级, FR_{ij} 为第 i 个因子下第 j 个等级的占比, w_{ij} 为第 i 个因子下第 j 个等级的滑坡数量, W 为区域滑坡总数, s_{ij} 为第 i 个因子下第 j 等级的栅格总数, S 为区域总栅格数. 若 $\text{FR}_{ij} \leq 0$, 表明因子对滑坡灾害发生影响较小; $\text{FR}_{ij} > 0$, 表明评价因子对滑坡灾害影响较大, 且 FR_{ij} 越大表明相关性越强.

归一化的计算公式如下:

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6)$$

式中, X 为因子被归一化的值, Y 为归一化后因子值,

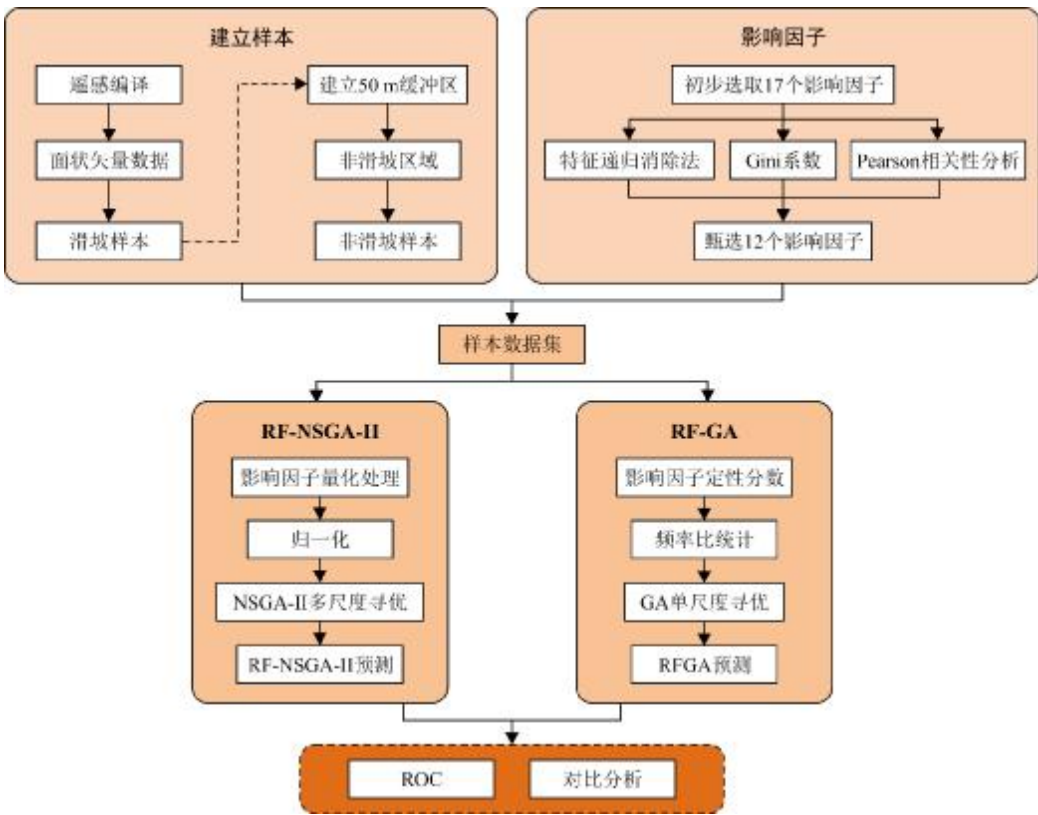


图 4 滑坡易发性评价流程图

Fig. 4 Flow chart for landslide susceptibility assessment



图 5 非滑坡样本选取图

Fig. 5 Selection of non-landslide samples

1—缓冲区域(buffer zone); 2—滑坡区域(landslide area); 3—非滑坡选取区域(selected area of non-landslide sample)

$X_{\min}$ 为样本中因子最小值, $X_{\max}$ 为样本中因子最大值.

3.5 NSGA-II 超参数优化

遗传算法来源于自然选择和遗传学的基本原理. 其关键流程包括个体初始化、适应度评估、选择操作、

交叉操作、变异操作和新一代的生成. 每个步骤都旨在逐步提高种群的整体适应度, 最终收敛到一个或多个最优解.

NSGA-II (非支配排序遗传算法 II) 是基于快速非支配排序技术的多目标优化算法, 通过帕累托支配和最优性概念对候选解进行有效分类^[26]. 该算法的核心在于使用非劣解分层来引导策略接近帕累托最优解集. NSGA-II 采用一种精英策略, 即在每一代中, 将父代与子代合并后进行非支配排序. 这一过程不仅扩大了搜索空间, 还允许更优的个体在下一代中有更高的生存几率. 在个体的选择过程中, 算法会优先考虑等级较高的个体, 而在同级个体中, 通过拥挤度进行选择, 从而保持种群多样性并确保优秀个体保留. 具体算法流程见图 6.

4 滑坡易发性评价

4.1 建立样本

4.1.1 滑坡-非滑坡样本

藏东南地区因其独特的构造背景条件、复杂的地

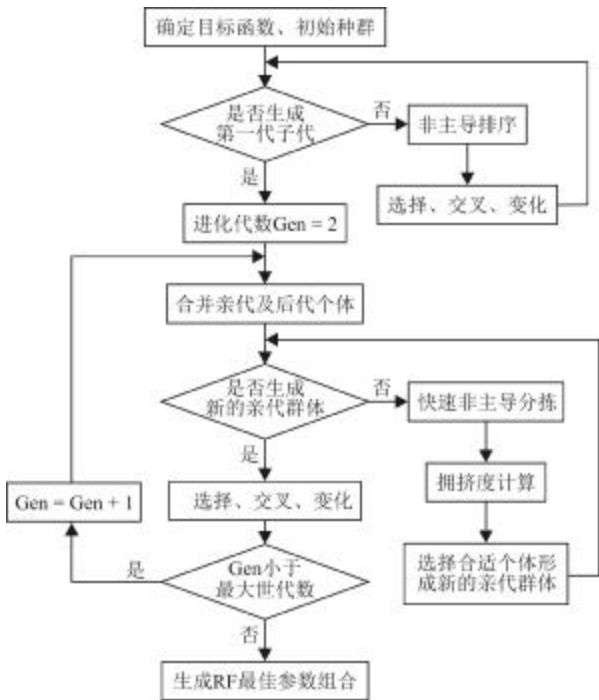


图 6 NSGA-II 算法流程图

Fig. 6 Flow chart of NSGA-II

形地貌以及多样的气候分区,地质灾害频发,表现出明显的区域性和地形特征.从空间和时间分布来看,该地区的滑坡灾害显著受到地质地貌背景和气候条件的控制.易发生滑坡的区域通常为河谷深切、山势陡峻、相对高差大的地区,特别是高山和极高山地区.为了确定非滑坡区域,以研究区滑坡面为基础建立 50 m 缓冲区,并在非滑坡区域内随机生成等量的非滑坡点,结合滑坡数据,最终组成 11 562 个滑坡-非滑坡样本数据,用于统计计算和模型训练.

4.1.2 影响因子量化处理及相关性分析

初步选取 17 个因子进行统一连续性及归一化处理,降低离散型的同时,使所有因子归于同一数据维度,保证机器学习的“学习效率”.各影响因子量化结果见表 2.

在滑坡易发性评价模型中,致灾因子的选取及重要性排序会对模型评估产生重要的影响^[27-28].本研究通过结合 Pearson 相关系数、随机森林 Gini 系数、特征递归消除法来评价指标之间的相关性并对因子重要性

表 2 主控因子量化结果表

Table 2 Quantitative results of main control factors

序号	类型	分类	影响因子	分级范围	归一化值
A	连续型	地形地貌因子	坡度	0~87°	0~1
B			坡向	0~360°	
C			高程	742.039~7109.46 m	
D			平面曲率	0~70.19	
E			剖面曲率	0~38.96	
F			地形起伏度	0~2294 m	
G			坡度变率	0~38.95	
H			坡向变率	0~70.19	
I	连续型	水文因子	距河流距离	0~5600 m	0~1
J	连续型	人类活动因子	距道路距离	0~4800 m	0~1
K	连续型	气象因子	多年平均降雨量	0~1289.88 mm	0~1
L	连续型	地质构造因子	距断层距离	0~40000 m	0~1
M	连续型	环境地质因子	地表切割深度	0~586.3	0~1
N			地形粗糙度	1~3.526	
O			归一化植被指数	-0.20~0.99	
P			土地利用类型	1~10 类	
Q	离散型	地层岩性因子	地层岩性分区	1~4 类	0~1

进行排序,对初选的 17 个影响因子进行筛选,由图 7 发现,地形起伏度与地表切割深度的 Pearson 相关性较大,系数为-0.87,因此需要剔除其中一个因子.基于随机森林 Gini 系数对各因子对滑坡贡献度的特征重要性分析(图 8)结果显示,地表切割深度的重要性较低,故予以剔除.此外,地形粗糙度的特征重要性为 0,即对滑坡发育坡向变率没有贡献,因此也被剔除.最后,采用特征递归消除算法再对 15 个影响因子进行剔除,最终甄选出地形地貌因子(坡度、坡向、高程、平面曲率、地形起伏度、坡度变率),人类活动因子(距道路距离),水文因子(距河流距离),地质构造因子(距断层距离),气象因子(多年平均降雨量),环境地质因子(土地利用类型)和地层岩性因子(地层岩性分区)12 个影响因子参与研究区的滑坡易发性评价.从每次剔除因子后模型 AUC 值变化可见(表 3),随着次要因子剔除,模型测试集 AUC 值不断提高,最终模型测试集 AUC

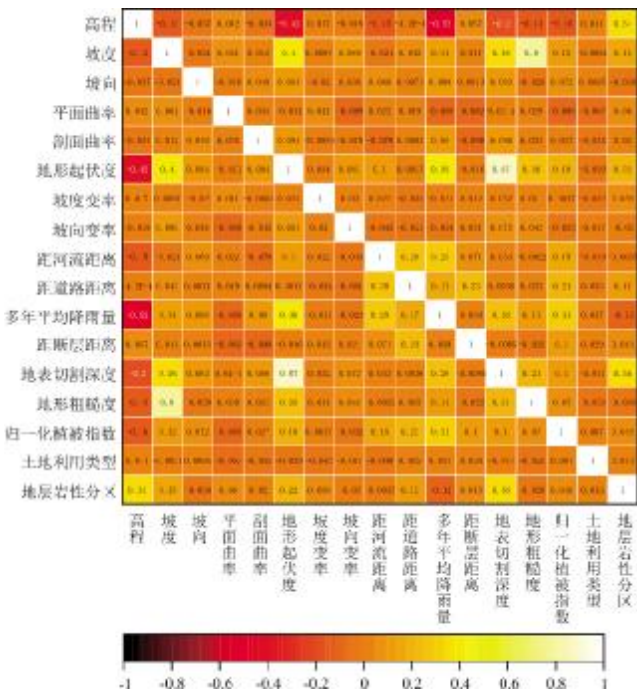


图 7 Pearson 相关系数热力图

Fig. 7 Heat map of Pearson correlation coefficients



图 8 主控因子特征重要性

Fig. 8 Importance of main control factor characteristics

值较未剔除因子提高了 0.032,模型精度得到了提升(图 9).

4.2 RF-NSGA-II 模型训练结果

基于 NSGA-II 的模型超参数自适应寻优显著改变了训练过程的时间成本.与默认参数设置相比,经过优化后的训练时长显著增加至 34 个小时.在优化过程中,尽管决策树数量减少,但其他寻优参数的数量却增加了.表明通过 NSGA-II 算法进行超参数调整,模型的训练时间虽延长,但也使得模型在更复杂的参数空间内找到了更优的配置,从而提高了模型的性能和预测能力(表 4).

在滑坡易发性评价模型中,经过优化后,距河流距离的重要性显著提升,从优化前的第四位上升至第三位,并且其在影响因子中的比重由 8.123%增加到 9.655%.这表明在滑坡发生的评估中,距离河流的远

表 3 测试集 AUC 值变化表

Table 3 AUC value changes of test set

影响因子筛选	未剔除	第 1 次剔除	第 2 次剔除	第 3 次剔除	第 4 次剔除	第 5 次剔除
剔除因子		地表切割深度	地形粗糙度	坡向变率	剖面曲率	NDVI
测试集 AUC 值	0.746	0.754	0.762	0.768	0.775	0.778

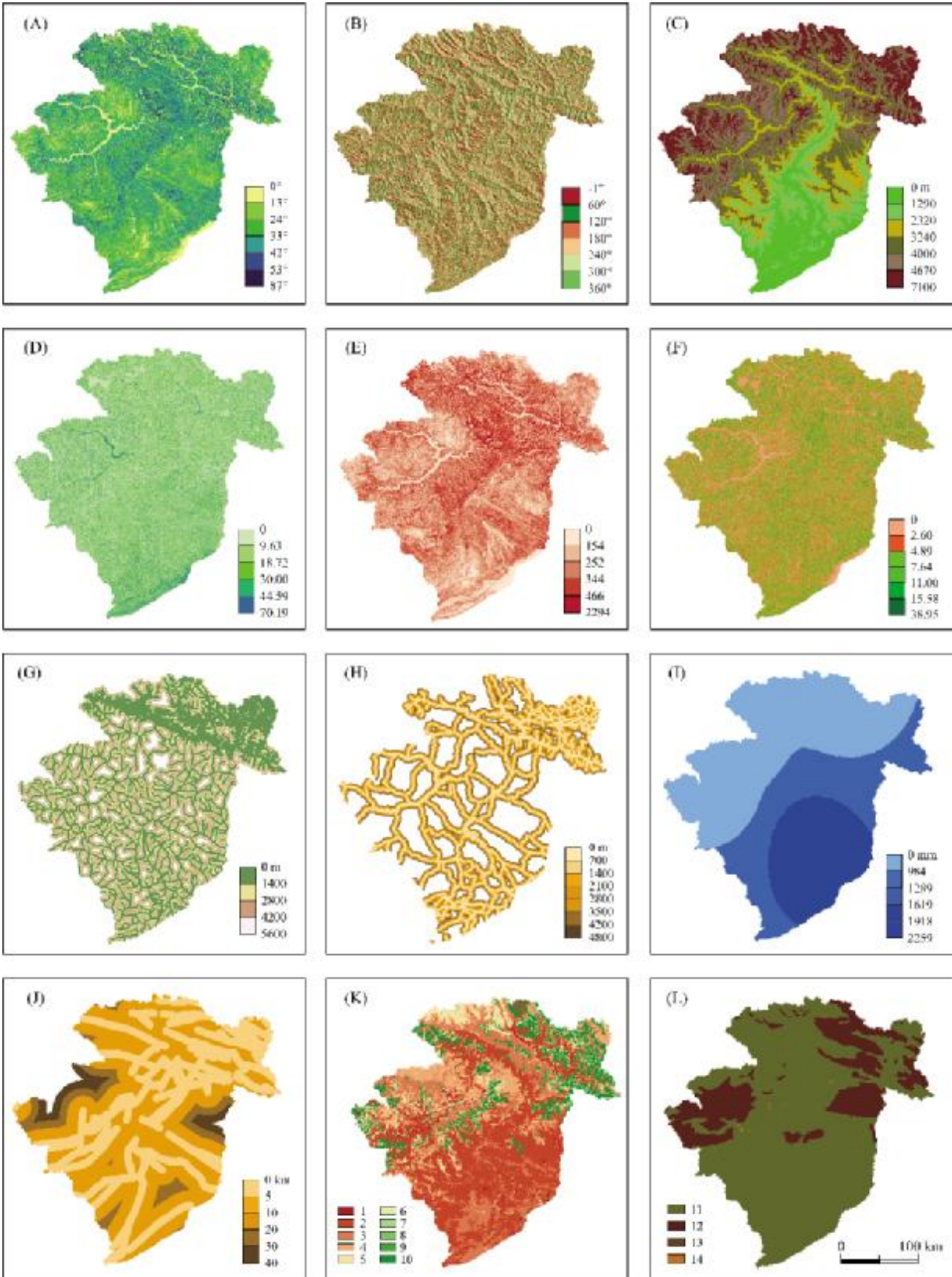


图 9 滑坡评价主控因子图

Fig. 9 Main control factor maps of landslide evaluation

A—坡度(slope gradient); B—坡向(slope aspect); C—高程(elevation); D—平面曲率(plane curvature); E—地形起伏度(topographic relief); F—坡度变率(slope variability); G—距河流距离(distance from river); H—距道路距离(distance from road); I—年平均降雨量(average annual rainfall); J—距断层距离(distance from fault); K—土地利用变率(land use variability); L—岩性分区(lithological zoning); 1—中覆盖林地(moderate-coverage forest land); 2—低覆盖林地(low-coverage forest land); 3—高覆盖林地(high-coverage forest land); 4—戈壁(gobi); 5—有林地(forest land); 6—河流、湖泊、水塘(river/lake/pond); 7—灌木林(shrub forest); 8—裸土(bare soil); 9—裸岩(bare rock); 10—其他未利用地(other unutilized land); 11—坚硬岩(hard rock); 12—软岩(soft rock); 13—较坚硬岩(mild hard rock); 14—较软岩(mild soft rock)

表 4 优化前后超参数对比
Table 4 Comparison of hyperparameters before and after optimization

参数名	RF-NSGA-II	RF-GA
训练用时	34 h 42 min 48 s	2s
树的数量	50	100
内部节点再分裂最小样本数	50	2
叶子节点所需最小样本数	48	1
树的最大深度	12	10

近对于预测和理解滑坡风险具有更大的影响力. 同时, 基于 Gini 系数的影响因子特征重要性在优化前后并未显示明显的变化. 这意味着除了距河流距离外, 其他影响因子在模型优化过程中对于滑坡贡献度的评估保持了相对稳定的状态(图 10).

在滑坡易发性评价模型中, 随机森林模型经过 NSGA-II 优化后, 训练精确率明显提升. F1 分数作为召回率与精确率的调和平均, 对评估模型的综合表现至关重要. 经过对比分析, 经过 NSGA-II 和 GA 优化前后的随机森林模型在训练结果中均表现出较高的准确性, 但 NSGA-II 优化后的模型训练精度由 0.794 提

升至 0.827, 显示出更高的准确性水平(表 5). 这表明通过使用 RF-NSGA-II 进行模型超参数的自适应寻优, 获得的最优指标组合为准确率(82.7%)、召回率(81.9%)、精确率(80.1%)、F1(81.5%).

表 5 优化前后训练集结果对比
Table 5 Comparison of training set results before and after optimization

训练集	准确率	召回率	精确率	F1
NSGA-II-RF	0.827	0.819	0.801	0.815
GA-RF	0.794	0.732	0.769	0.796

4.3 滑坡易发性评价结果

使用训练好的模型进行研究区域的滑坡易发性评估, 得到 RF-NSGA-II 和 RF-GA 滑坡易发性评价结果(图 11). 通过自然间断法将研究区域划分为高、较高、中、较低、低 5 个易发性等级, 并计算了各等级区域的面积以及各区域内的滑坡面积(表 6). 与 RF-GA 预测结果进行对比分析后发现, RF-NSGA-II 预测结果显示出高、较高、中 3 个易发性等级面积占比逐渐递增. 特别是在高、中易发区域的滑坡分布更为集中, 相比于 RF-GA 预测结果展现出明显的规律性和合理性.

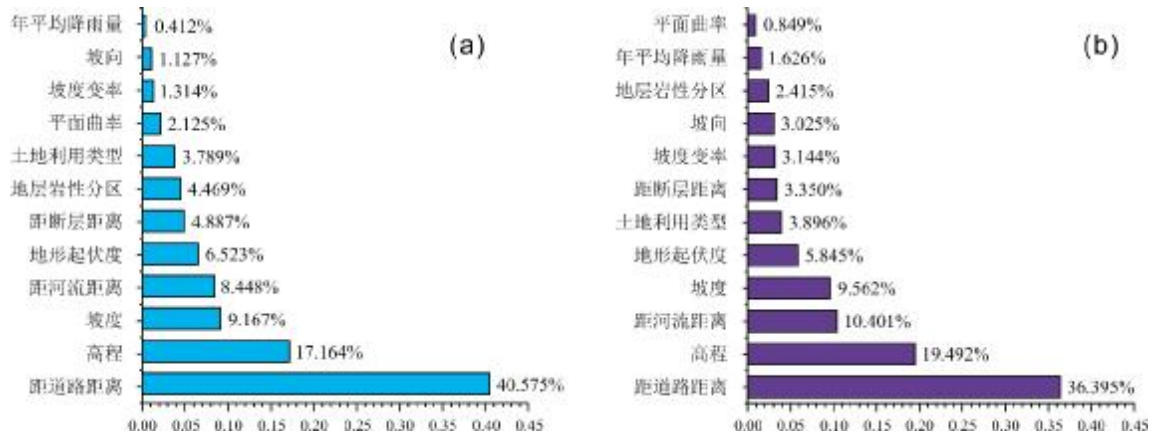


图 10 RF-NSGA-II 和 RF-GA 模型的主控因子特征重要性对比图
Fig. 10 Importance comparison of the main control factor characteristic between RF-NSGA-II and RF-G models
a— RF-NSGA-II 模型(RF-NSGA-II model); b— RF-GA 模型(RF-GA model)

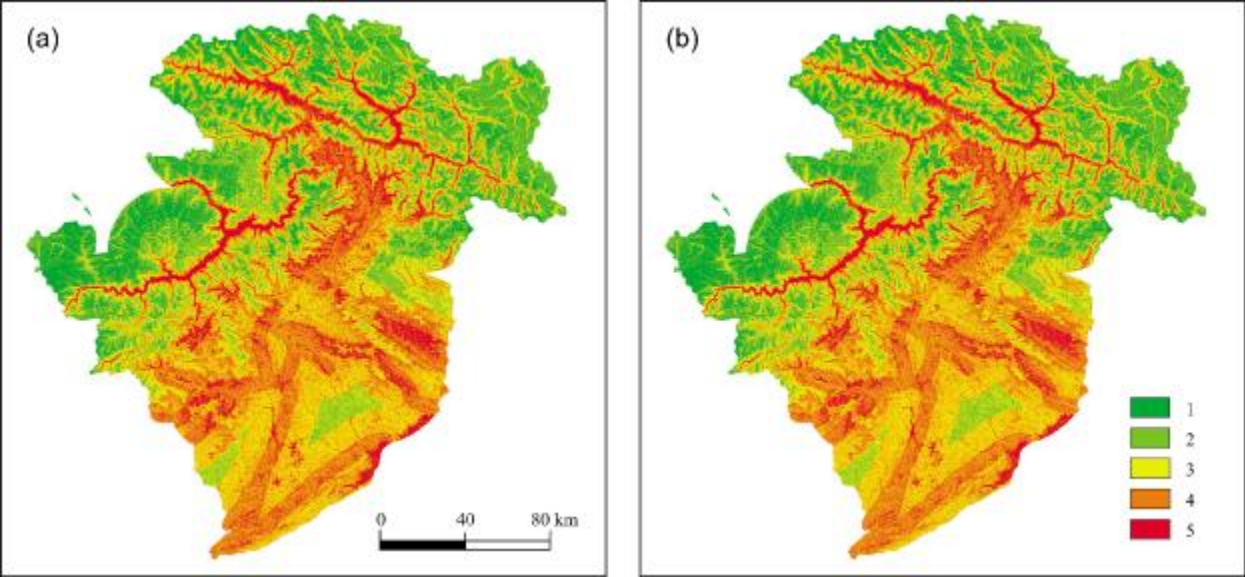


图 11 滑坡易发性评价结果对比图

Fig. 11 Comparison of landslide susceptibility assessment results
a— RF-NSGA-II 模型(RF-NSGA-II model); b— RF-GA 模型(RF-GA model); 1—低易发性(lower susceptibility); 2—较低易发性(low susceptibility); 3—中易发性(medium susceptibility); 4—较高易发性(high susceptibility); 5—高易发性(higher susceptibility)

表 6 滑坡易发性评价分区统计结果

Table 6 Statistics of landslide susceptibility assessment zoning

易发性等级	高		较高		中		较低		低	
	面积占比/%	分区面积	滑坡面积	分区面积	滑坡面积	分区面积	滑坡面积	分区面积	滑坡面积	分区面积
RF-NSGA-II		44.42	14.95	28.79	29.52	15.46	4.63	2.34	17.94	2.34
RF-GA		41.69	12.25	28.66	26.52	21.67	23.63	8.71	21.94	2.68

4.4 ROC 曲线验证

AUC(ROC 曲线下的面积)是衡量模型泛化性能的重要指标^[29]. 根据 ROC 曲线的展示,模型在滑坡易发性预测方面表现出较高的准确性. RF-NSGA-II 模型显示出更高的 AUC 值,达到了 0.877,明显优于 RF-GA 模型预测结果(图 12).

本研究通过使用相同的滑坡-非滑坡输入样本,在滑坡易发性评价中对比了 RF-NSGA-II 模型和 RF-GA 模型的评价结果. RF-NSGA-II 模型和 RF-GA 模型在建模过程中使用相同的输入数据,保证了评价结果的精度和可比性. RF-GA 模型已经达到了较高的准确率水平. 然而,通过在因子定量特征和超参数设置上进一步优化和完善,获得最优指标组合,RF-NSGA-II 模型显示出更高的预测性能.

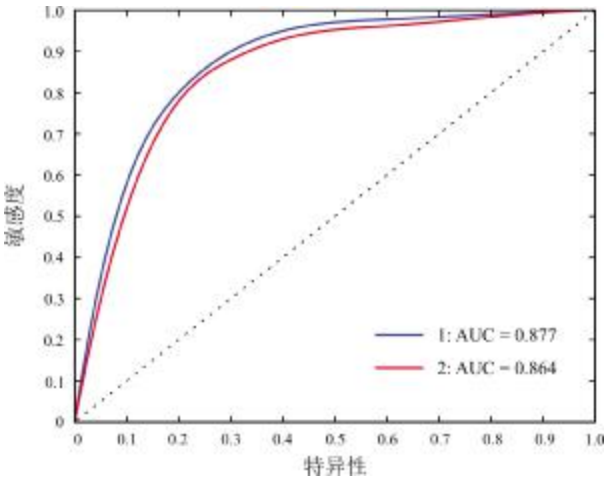


图 12 滑坡易发性评价结果 ROC 曲线

Fig. 12 ROC curves of landslide susceptibility assessment results
1— RF-NSGA-II 模型评价结果(RF-NSGA-II model); 2—RF-GA 模型评价结果(RF-GA model)

4.5 结果验证与分析

4.5.1 典型实例

为验证模型在滑坡易发性评估中的准确性和适用性,选择了具有河谷深切、山势陡峻、相对高差大等地理特征的雅鲁藏布江中下游及其支流作为重点研究区域.在这些条件特殊的区域内,详细调查和分析了多处滑坡.具体而言,选取了江北公路附近、派镇直白村 I 号滑坡、雅鲁藏布江流域的米林村滑坡、易贡藏布流域的 K47 滑坡以及 H03 滑坡等典型案例进行研究.通过计算高、较高、中、较低、低 5 个分区易发性面积占比,与 RF-GA 模型进行对比,RF-NSGA-II 模型预测结果中高易发性占比面积由 7.98% 提升至 9.32%,较高易发性占比面积由 14.12% 提升至 15.85%,且流域附近典型滑坡点所处区域均为高易发区(图 13).

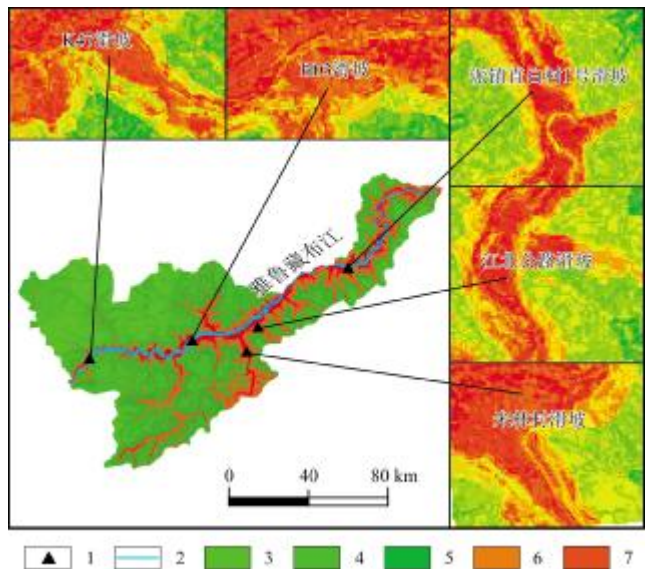


图 13 RF-NSGA-II 模型的流域附近滑坡易发性结果

Fig. 13 Susceptibility assessment results of landslide near watershed by RF-NSGA-II model

1—典型滑坡点(typical landslide site); 2—水系(water system); 3—低易发性(lower vulnerability); 4—较低易发性(low susceptibility); 5—中易发性(medium susceptibility); 6—较高易发性(high susceptibility); 7—高易发性(higher susceptibility)

1)江北公路滑坡

选取森波-帮羌纳乡河段附近的江北公路滑坡,由于特殊的地质和地形条件,该滑坡点平均坡度达 55°,投影面积为 0.45 km²,平均厚度为 165 m,总体积达到 1.5×10⁸ m³.该地区滑坡周围分布着 3 条断层,断

层密度为 63.125 km/100 km²,是区域平均密度的 2.85 倍.巴嘎沟位于嘉黎深切断裂带和雅鲁藏布江断裂带附近,地质构造活跃,岩石组成多样,包括中厚层砂岩、片岩夹灰岩及砂岩岩组.持续的构造作用,特别是嘉黎深切断裂带和雅鲁藏布江断裂带的影响,使岩石特别易于破碎,增加了滑坡发生可能性.由模型对比分析(表 7),此河段预测出高易发区面积从 8.16% 增加到 9.18%,RF-NSGA-II 模型易发性分区结果与实际滑坡区较为一致,且优于 RF-GA 模型预测结果.

表 7 典型滑坡群样本分布表

Table 7 Sample distribution of typical landslide groups

分段	河道长/km	滑坡数/处	高易发区面积占比(优化 RF)/%	高易发区面积占比(常规 RF)/%	滑坡密度/(处/km²)
派镇麦朗村-直白村	16.8	23	5.87	4.96	1.24
巴嘎沟-沟东	22.6	28	7.14	7.64	1.16
索松村-蹦嘎	23.8	27	6.89	6.23	0.87
本宗-增多	34.9	32	8.16	7.89	0.68
丹娘乡-卧龙镇	77.2	48	16.67	15.23	0.63
卧龙-米堆	57.8	44	12.24	10.28	0.57
森波-帮羌纳乡	52.8	36	9.18	8.16	0.48
玉松村-日加贡巴	16.8	15	3.83	2.96	0.41
朗多-白拉村	92.3	45	11.48	10.48	0.26
岗派公路附近	87.9	32	8.16	7.96	0.25
派村-格嘎村	91.7	15	3.83	3.21	0.19
羌纳乡-则隆弄	58.4	12	3.06	2.89	0.14
米堆-101 道班	284.5	35	8.93	8.36	0.11

2)派镇直白村 I 号滑坡

选取派镇麦朗村至直白村河段附近的直白村 I 号滑坡,前缘海拔最低 2 120 m,后缘最高 2 648 m,相对高差达 528 m.其体长 530 m,宽 320 m,平均坡度为 56°,总面积 3.65 km²,体积 5.15×10⁴ m³.该区域处于板块构造的交汇处,周围存在两条明显的断裂带,断层密度为 50.487 km/100 km²,是平均密度的 2.48 倍.地形特征显示,高海拔平台与河床之间的陡坡上存在古老的冰碛平台,高差达 456 m,使得斜坡区域的物质具有

显著的势能. 大部分坡体坡度超过 35° , 表明存在大量地形凸起体, 总量达 $0.196 \times 10^8 \text{ m}^3$. 这些凸起体为滑坡提供了物质来源, 而陡峭的边坡则为滑坡的发生提供了势能条件. 由模型对比分析(表 7), 此河段预测出高易发区面积从 4.96% 增加到 5.87%, RF-NSGA-II 模型易发性分区结果与实际滑坡区较为一致, 且优于 RF-GA 模型预测结果.

3) 米林村滑坡

选取卧龙镇至米堆村河段附近的米林村滑坡, 海拔从最低点 962 m 到最高点 1 350 m, 相对高差达 388 m. 其平均坡度约为 34° , 滑坡面积约为 $21.02 \times 10^4 \text{ m}^2$, 体积估计为 $415.8 \times 10^4 \text{ m}^3$. 在地质构造上, 该滑坡被米林次级断裂带切割, 邻近还有两条断裂带, 断层密度达到 $43.56 \text{ km}/100 \text{ km}^2$, 是平均密度的 2.51 倍. 滑坡中部地形凸起明显, 这些特征为其滑动提供了势能条件. 当前情况显示, 滑坡仅表层发生了位移, 后缘出现明显的裂缝, 若整体滑动, 可能导致河道阻塞. 由模型对比分析(表 7), 此河段预测出高易发区面积从 10.28% 增加到 12.24%, RF-NSGA-II 模型易发性分区结果与实际滑坡区较为一致, 且优于 RF-GA 模型预测结果.

4) K47 滑坡

选取岗派公路附近的 K47 滑坡, 该区域滑坡前缘海拔最低约 2 615 m, 后缘海拔最高达 2 956 m, 高差为 341 m. 滑坡面积为 $6.08 \times 10^4 \text{ m}^2$, 平均厚度约 21 m, 平均坡度为 53° . 据估算, 滑坡的体积约为 $102.3 \times 10^4 \text{ m}^3$. 初步判断, 该滑坡属于浅层牵引式滑坡类型. 在地质构造方面, 滑坡周围存在 3 条断层, 雅鲁藏布江断裂带贯穿滑坡区域, 断层密度为 $67.45 \text{ km}/100 \text{ km}^2$, 是平均密度的 3.56 倍. 受断裂带影响, 滑坡区域的基岩主要由坚硬到较坚硬的中厚层砂岩夹泥岩和板岩组成. 现场调查显示, 在滑坡发生之前, 滑坡上存在许多凸起的地形特征. 这些凸起的地形特征为滑坡的物质来源提供了条件, 而陡峭的边坡则为滑坡的发生提供了势能条件. 由模型对比分析(表 7), 此河段预测出高易发区面积从 7.96% 增加到 8.16%, RF-NSGA-II 模型易发性分区结果与实际滑坡区较为一致, 且优于 RF-GA 模型预测结果.

5) H03 滑坡

选取派村-格嘎村附近的 H03 滑坡, 该区域前缘

最低海拔为 3 021 m, 后缘最高海拔为 3 225 m, 相对高差为 204 m. 滑坡面积约为 $8.21 \times 10^4 \text{ m}^2$, 长度为 325 m, 平均坡度为 37° , 体积估计为 $123.7 \times 10^4 \text{ m}^3$, 属于大型滑坡. 在地质构造方面, 雅鲁藏布江断裂带贯穿了滑坡区域, 相对断层密度为 $43.56 \text{ km}/100 \text{ km}^2$, 是平均密度的 2.56 倍. 受到断裂带的影响, 滑坡后缘发生了垮塌, 形成地貌堆积体, 滑坡体上出现明显的裂缝, 稳定性较差. 滑坡中部有道路穿越, 前缘受到河流的冲刷, 这些因素为滑坡的发生提供了势能和动能条件. 由模型对比分析(表 7), 此河段预测出高易发区面积从 3.21% 增加到 3.83%, RF-NSGA-II 模型易发性分区结果与实际滑坡区较为一致, 且优于 RF-GA 模型预测结果.

4.5.2 滑坡发育与主控因素分析

为了深入探究区域滑坡灾害的发育与主控因素, 本研究将区域内河道细分为 11 段, 并通过对灾害发育密度的统计, 系统分析区域滑坡灾害的主要影响因素(表 7). 这一划分基于河道地貌条件和特征, 旨在捕捉沿线不同区域的滑坡活动情况, 从而更全面地理解滑坡灾害的形成机制和空间分布规律. 通过统计每个河段的高易发区面积占比及滑坡灾害发育密度, 能够准确评估不同区域的滑坡灾害程度及其对整个研究区的影响程度. 同时, 分析各段之间的差异可以揭示出可能的主控因素, 并进一步验证这些因素在滑坡灾害发育中的作用和重要性.

研究区域内共发育 392 处滑坡, 与 RF-GA 模型对比, RF-NSGA-II 模型预测出各个河段高易发区面积占比均有增加, 其中派镇麦朗村至直白村、巴嘎沟至结东、索松村至蹦蹦滑坡密度较大. 该区位于雅鲁藏布江大拐弯地区, 是易贡藏布和帕隆藏布江的交汇处, 也是高断层密度区域之一. 特别是在米林镇至加热萨乡段(雅鲁藏布江大拐弯), 分布着大型以上规模的滑坡, 这一区域位于喜马拉雅山脉、横断山脉和念青唐古拉山脉的交汇处, 是东喜马拉雅构造带北缘, 也是构造应力最为集中的地段. 受到强烈的构造应力作用影响, 大拐弯顶端作为隆升的核心区, 隆升造成了新的不稳定和不平衡, 是该地区滑坡发育的主要控制因素. 因此, 综合地质构造和地貌特征分析, 揭示了该区域滑坡灾害发生的复杂机制和空间分布规律, 为进一步的灾害预测和防范提供了重要依据.

5 讨论

本研究在藏东南部分区域进行了滑坡灾害易发性评价,采用 RF-NSGA-II 模型(多目标优化)与 RF-GA (单目标优化)进行对比,4 项指标值(准确度、召回率、精确度、F1)均有所提高。研究选取了滑坡和非滑坡样本集共计 12 个影响因子进行评估,最终模型的精度达到了 0.877。研究结果显示,RF-NSGA-II 模型高易发区、较高易发区、中易发区面积占比分别为 14.95%, 29.52%, 24.63%, 高于 RF-GA 模型。高易发滑坡区域主要集中在易贡藏布和帕隆藏布江的交汇处以及雅鲁藏布江大拐区域。这些区域长期受复杂的地质构造影响,东西两侧有巨型断裂,南侧为喜马拉雅前缘断裂带,北侧为冈底斯造山带。该地区地质构造复杂,地貌丰富,岩浆活动异常发育,为滑坡灾害频发提供了基础条件。

研究还表明,基于 RF-NSGA-II 模型超参数自适应寻优可以提升模型精度。尽管本次研究未对多种机器学习或深度学习算法超参数进行深入调试,但仍在提高训练效果上取得了一定效果。

随着海量数据积累和算力提升,机器学习在地质灾害评价中的应用前景广阔。针对本文研究的优势和局限性,未来的工作可以从以下几个方面展望:1)在滑坡-非滑坡样本数据处理上,优化非滑坡样本筛选方法以提高准确率;2)探索不同离散型因子的量化表征方式,以降低数据离散性,更好地适配机器学习模型;3)进一步优化其他机器学习模型的超参数自适应寻优算法,以在保持精度的前提下提升效率和稳定性。

6 结语

本研究选取了藏东南部分区域,针对滑坡-非滑坡样本筛选、影响因子选取、联接方法应用和超参数寻优 4 个方面,在滑坡易发性评价中使用 RF-NSGA-II 模型(多目标优化),主要得出以下结论:

1) 基于 11 562 处滑坡面状矢量数据,采用建立 50 m 缓冲区反向选取的方法,获取较为准确的非滑坡样本集,并结合 Pearson 相关性分析、Gini 系数、特征递归消除法,确定了距道路距离、坡度等 12 个影响因子。其中,距道路距离、高程和距河流距离分别位列前三,对研究区滑坡主控因子进行分析,高易发区主要分布在易贡藏布、帕隆藏布交汇处以及雅鲁藏布江大拐

弯区域。该地区地质构造复杂,断层纵横交错,岩石易于破碎,节理发育,同时受到频繁的构造活动影响。这些因素导致了地表的隆升和河流的强烈冲刷,形成了陡峭的坡度和高差,为滑坡的发生提供了丰富的能量条件。

2) 结合 NSGA-II(多目标优化)对随机森林模型进行了自适应寻优,经过 34 h 的模型训练,得出了决策树数量和树的最大深度等超参数的最佳配置,实现最优参数组合:准确度(0.827)、召回率(0.819)、精确度(0.801)、F1(0.827)。比较 RF-NSGA-II 模型与 RF-GA 模型的 ROC 曲线和 AUC 值,发现 RF-NSGA-II 模型的 AUC 值达到 0.877,显示出更高的预测精度。分析显示,87%以上的滑坡位于中易发区以上区域,其中高易发区的滑坡密度达到 15.6 处/100 km²,而低易发区的滑坡密度仅为 0.028 处/100 km²。这一结果表明易发区等级与滑坡密度之间存在正相关关系。

3) 为验证模型的性能,本研究选取了 2 处典型滑坡点和 3 个野外调查滑坡点进行验证。计算断层密度结果显示,滑坡点的断层密度均超过了 35 km/100 km²,明显高于研究区域的平均断层密度(18.64 km/100 km²)。这一结果揭示了这些区域地质构造环境的复杂性和活跃性,进一步确认了 RF-NSGA-II 模型结果的合理性和准确性。

参考文献(References):

- [1] 许强. 对滑坡监测预警相关问题的认识与思考[J]. 工程地质学报, 2020, 28(2): 360-374.
Xu Q. Understanding the landslide monitoring and early warning: Consideration to practical issues[J]. Journal of Engineering Geology, 2020, 28(2): 360-374.
- [2] 韩用顺. 藏东南地质灾害易发性评价[J]. 山地学报, 2021, 39(5): 700.
Han Y S. Assessment of geological disaster susceptibility in southeastern Tibet[J]. Mountain Research, 2021, 39(5): 700. (in Chinese)
- [3] 冉涛, 徐如阁, 周洪福, 等. 雅鲁江流域深切河谷区滑坡类型、成因及分布规律——以子拖西-麻郎错河段为例[J]. 中国地质, 2024, 51(2): 511-524.
Ran T, Xu R G, Zhou H F, et al. Type, formation mechanism and distribution regularity of landslides in the deeply incised valley area of Yalong River Basin: A case study of Zituoxi-Malangcuo River section [J]. Geology in China, 2024, 51(2): 511-524.
- [4] 牛犇, 冯春, 丛俊余, 等. 基于 CDEM 颗粒流的三维高速远程滑坡成灾范围分析[J]. 岩石力学与工程学报, 2023, 42(S2): 4018-4027.

- Niu B, Feng C, Cong J Y, et al. Three-dimensional high-speed remote landslide disaster area analysis based on CDEM particle flow [J]. Chinese Journal of Rock Mechanics and Engineering, 2023, 42 (S2): 4018–4027.
- [5] 袁浩, 郭长宝, 吴瑞安, 等. 西藏易贡高位远程滑坡研究进展与展望[J]. 地质通报, 2023, 42(10): 1757–1773.
- Yuan H, Guo C B, Wu R A, et al. Research progress and prospects of the giant Yigong long run-out landslide, Tibetan Plateau, China[J]. Geological Bulletin of China, 2023, 42(10): 1757–1773.
- [6] 王哲, 赵超英, 刘晓杰, 等. 西藏易贡滑坡演化光学遥感分析与 InSAR 形变监测[J]. 武汉大学学报(信息科学版), 2021, 46(10): 1569–1578.
- Wang Z, Zhao C Y, Liu X J, et al. Evolution analysis and deformation monitoring of Yigong landslide in Tibet with optical remote sensing and InSAR[J]. Geomatics and Information Science of Wuhan University, 2021, 46(10): 1569–1578.
- [7] 韩盼, 余国安, 侯伟鹏. 藏东南帕隆藏布流域冲积扇地貌发育特征及空间分异研究[J]. 地理科学进展, 2024, 43(4): 784–798.
- Han P, Yu G A, Hou W P. Geomorphological development characteristics and spatial differentiation of alluvial fans in high mountain canyon areas: A case study in the Parlung Tsangpo Basin, southeast Tibet[J]. Progress in Geography, 2024, 43(4): 784–798.
- [8] 杨帆, 马志刚, 文艳, 等. 联合 SAR 影像相位和幅度信息的多形变量级滑坡监测——以金沙江白格滑坡为例[J]. 自然资源遥感, 2024, 36(2): 257–267.
- Yang F, Ma Z G, Wen Y, et al. Monitoring a landslide with a multi-deformation magnitude based on the phase and amplitude information of SAR images: A case study of the Baige landslide in Jinsha River [J]. Remote Sensing for Natural Resources, 2024, 36(2): 257–267.
- [9] 付杰, 张青, 仇文岗. 基于多源数据与机器学习的降雨型滑坡灾害危险性动态预警[J]. 地质论评, 2024, 70(S1): 217–218.
- Fu J, Zhang Q, Zhang W G. Dynamic warning of rainfall-induced landslide hazard based on multi-source data and machine learning[J]. Geological Review, 2024, 70(S1): 217–218.
- [10] 张冬梅, 张莉. 对话姚檀栋: 走近第二次青藏高原综合科学考察[J]. 科学通报, 2019, 64(27): 2765–2769.
- Zhang D M, Zhang L. Tandong Yao: The second Tibetan Plateau scientific expedition and research [J]. Chinese Science Bulletin, 2019, 64(27): 2765–2769.
- [11] 王悦, 曹颖, 许方党, 等. 考虑非滑坡样本选取和集成机器学习方法的水库滑坡易发性预测[J]. 地球科学, 2024, 49(5): 1619–1635.
- Wang Y, Cao Y, Xu F D, et al. Reservoir landslide susceptibility prediction considering non-landslide sampling and ensemble machine learning methods[J]. Earth Science, 2024, 49(5): 1619–1635.
- [12] 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.
- [13] 茹曼, 郑燕, 张斌, 等. 基于 SVM-RF 模型的地质灾害易发性评价——以河南省博爱县青天河景区为例[J]. 地质与资源, 2023, 32(5): 633–641. DOI: 10.13686/j.cnki.dzyzy.2023.05.014.
- Ru M, Zheng Y, Zhang B, et al. Assessment of geological disaster susceptibility based on SVM-RF MODEL: A case study of Qingtian River scenic area in Boai County, Henan Province[J]. Geology and Resources, 2023, 32(5): 633–641. DOI: 10.13686/j.cnki.dzyzy.2023.05.014.
- [14] 李成林, 刘严松, 赖思翰, 等. 基于逻辑回归和支持向量机耦合模型的滑坡易发性分析[J]. 自然灾害学报, 2024, 33(2): 75–86.
- Li C L, Liu Y S, Lai S H, et al. Landslide susceptibility analysis based on the coupling model of logistic regression and support vector machine[J]. Journal of Natural Disasters, 2024, 33(2): 75–86.
- [15] 杜鹏, 陈宁生, 伍康林, 等. 基于随机森林模型的藏东南地区滑坡易发性评价及主控因素分析[J]. 成都理工大学学报(自然科学版), 2024, 51(2): 328–344.
- Du P, Chen N S, Wu K L, et al. Evaluation of landslide susceptibility in southeast Tibet based on a random forest model [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2024, 51(2): 328–344.
- [16] 林泊锟, 丁勇, 李登华. 一种基于深度学习与图像局部特征提取的边坡异常监测技术[J]. 测绘通报, 2024(4): 23–28.
- Lin B K, Ding Y, Li D H. A slope anomaly monitoring technology based on deep learning and image local feature extraction[J]. Bulletin of Surveying and Mapping, 2024(4): 23–28.
- [17] 冯旻譞, 毛伊敏, 贾俊, 等. 地质灾害危险性评价中不同机器学习方法优劣对比: 以宁强县大安镇为例[J/OL]. 中国地质, 1–13 [2024–06–17].
- Feng M X, Mao Y M, Jia J, et al. Comparison of the advantages and disadvantages of different machine learning methods in geohazard risk assessment: Taking Da'an Town, Ningqiang County as an example [J/OL]. China Geology, 1–13 [2024–06–17].
- [18] 蒋宏伟, 刘健鹏, 王新杰, 等. 随机森林优化的静态耦合模型在滑坡位移预测中的应用[J]. 常州大学学报(自然科学版), 2024, 36(3): 80–92.
- Jiang H W, Liu J P, Wang X J, et al. Application of static dynamic coupling model optimized by random forest in landslide displacement prediction [J]. Journal of Changzhou University (Natural Science Edition), 2024, 36(3): 80–92.
- [19] 黄发明, 欧阳慰平, 蒋水华, 等. 考虑机器学习建模中训练/测试集时空划分原则的滑坡易发性预测建模[J]. 地球科学, 2024, 49(5): 1607–1618.
- Huang F M, Ouyang W P, Jiang S H, et al. Landslide susceptibility

- prediction considering spatio-temporal division principle of training/testing datasets in machine learning models[J]. *Earth Science*, 2024, 49(5): 1607–1618.
- [20] 王本栋, 李四全, 许万忠, 等. 基于3种不同机器学习算法的滑坡易发性评价对比研究[J]. *西北地质*, 2024, 57(1): 34–43.
- Wang B D, Li S Q, Xu W Z, et al. A comparative study of landslide susceptibility evaluation based on three different machine learning algorithms[J]. *Northwestern Geology*, 2024, 57(1): 34–43.
- [21] 周剑, 汤明高, 裴芳歌, 等. 基于机器学习的库岸滑坡变形短期预测[J]. *山地学报*, 2023, 41(6): 891–903.
- Zhou J, Tang M G, Pei F G, et al. Short-term deformation of reservoir slope based on machine learning[J]. *Mountain Research*, 2023, 41(6): 891–903.
- [22] 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. Geological disaster susceptibility evaluation based on coupling of frequency ratio model and random forest model[J]. *Journal of Natural Disasters*, 2023, 32(6): 74–82.
- [23] Merghadi A, Yunus A P, Dou J, et al. Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance[J]. *Earth-Science Reviews*, 2020, 207: 103225.
- [24] Pawluszek K, Borkowski A. Impact of DEM-derived factors and analytical hierarchy process on landslide susceptibility mapping in the region of Rożnów Lake, Poland[J]. *Natural Hazards*, 2017, 86(2): 919–952.
- [25] Yang H, Shi, P J, Quincey D, et al. A heterogeneous sampling strategy to model earthquake-triggered landslides[J]. *International Journal of Disaster Risk Science*, 2023, 14(4): 636–648.
- [26] Tabatabaei M, Salehpour Jam A, Ahmad Hosseini S. Suspended sediment load prediction using non-dominated sorting genetic algorithm II [J]. *International Soil and Water Conservation Research*, 2019, 7(2): 119–129.
- [27] 吴辰文, 梁靖涵, 王伟, 等. 基于递归特征消除方法的随机森林算法[J]. *统计与决策*, 2017(21): 60–63.
- Wu C W, Liang J H, Wang W, et al. Random forest algorithm based on recursive feature elimination[J]. *Statistics & Decision*, 2017(21): 60–63.
- [28] 孙德亮, 陈丹璐, 密长林, 等. 基于随机森林-特征递归消除模型的可解释性缓丘岭谷地貌滑坡易发性评价[J]. *地质力学学报*, 2023, 29(2): 202–219.
- Sun D L, Chen D L, Mi C L, et al. Evaluation of landslide susceptibility in the gentle hill-valley areas based on the interpretable random forest-recursive feature elimination model[J]. *Journal of Geomechanics*, 2023, 29(2): 202–219.
- [29] Cantarino I, Carrion M A, Goerlich F, et al. A ROC analysis-based classification method for landslide susceptibility maps[J]. *Landslides*, 2019, 16(2): 265–282.

(上接第43页/Continued from Page 43)

- [22] 刘驰, 张华, 汤正江, 等. 我国森林沼泽景观区地球化学系列参数统计[J]. *物探与化探*, 2013, 37(4): 585–590.
- Liu C, Zhang H, Tang Z J, et al. Statistics of series of geochemical parameters for the forest swamp landscape in China[J]. *Geophysical and Geochemical Exploration*, 2013, 37(4): 585–590.
- [23] 柳坤峰, 冯昌荣, 翟黎明, 等. 新疆乌恰县吾合沙鲁地区水系沉积物地球化学特征与找矿远景[J]. *现代地质*, 2019, 33(4): 759–771.
- Liu K F, Feng C R, Zhai L M, et al. Geochemical characteristics of stream sediments and prospecting direction in the Wuheshalu area, Wujia County, Xinjiang[J]. *Geoscience*, 2019, 33(4): 759–771.
- [24] 胡以铿. 地球化学中的多元分析[M]. 武汉: 中国地质大学出版社, 1991: 1–229.
- Hu Y K. Multivariate analysis in geochemistry [M]. Wuhan: China University of Geosciences Press, 1991: 1–229. (in Chinese)