

骆飞, 凌斯祥, 高凤欣, 等. 基于不同滑坡负样本选取策略和随机森林方法的岷江上游大型滑坡易发性评价[J]. 水利水电技术(中英文), 2025, 56(9): 42-59. DOI: 10.13928/j.cnki.wrahe.2025.09.004

LUO Fei, LING Sixiang, GAO Fengxin, et al. Large-scale landslide susceptibility evaluation in upper reaches of Minjiang River based on different selection methods for landslide negative samples and random forest algorithm[J]. Water Resources and Hydropower Engineering, 2025, 56(9): 42-59. DOI: 10.13928/j.cnki.wrahe.2025.09.004

基于不同滑坡负样本选取策略和随机森林方法的岷江上游大型滑坡易发性评价

骆飞¹, 凌斯祥^{1,2}, 高凤欣³, 林祖豪¹, 孙春卫⁴, 高芳芳⁵, 巫锡勇^{1,2}

(1. 西南交通大学 地球科学与工程学院, 四川 成都 611756; 2. 四川省环青藏高原交通廊道地质灾害生态化防治工程技术研究中心, 四川 成都 611756; 3. 西藏大学 工学院, 西藏 拉萨 850000; 4. 西华大学 应急管理学院, 四川 成都 610065; 5. 成都锦城学院, 四川 成都 611731)

摘要:【目的】依据准确的滑坡易发性结果能对滑坡危险性与风险性进行精准防控。在滑坡易发性评价中,不同滑坡负样本的选取方式是影响滑坡易发性预测精度的重要不确定因素。【方法】以四川岷江流域山区为研究区,通过遥感影像编制881个大型滑坡($> 10^6 \text{ m}^3$),选取地形地貌、基础地质、水文地质、地质环境、地震参数与人类活动等13个滑坡评价因子,并共线性分析检验因子冗余;其次对滑坡负样本采取全区随机、坡度低于 10° 区域、滑坡缓冲1 km外区域、信息量法(IV)、支持向量机法(SVM)和半监督法选择出与滑坡等比例的滑坡负样本,进一步与随机森林(RF)耦合构建随机RF、低坡度RF、缓冲区RF、IV-RF、SVM-RF与半监督RF模型开展滑坡易发性区划;最后采用ROC曲线的平均AUC值对不同滑坡负样本采样模型的预测精度进行对比评价。【结果】结果显示:(1)不同滑坡负样本采样方式得到的滑坡高易发区与极高易发区均集中分布于卧龙至映秀段、棉鹿至古尔沟段、黑虎至木苏段与叠溪至松潘段河谷两侧;(2)不同滑坡负样本采样方式的滑坡易发性预测精度为:半监督RF($\overline{AUC}=0.971$)>SVM-RF($\overline{AUC}=0.954$)>IV-RF($\overline{AUC}=0.945$)>缓冲区RF($\overline{AUC}=0.902$)>低坡度RF($\overline{AUC}=0.895$)>随机RF($\overline{AUC}=0.882$);(3)在低易发区选择滑坡负样本能明显提高易发性精度,低坡度RF、缓冲区RF、IV-RF、SVM-RF与半监督RF模型相较于随机RF模型,AUC值分别提高了0.013、0.02、0.063、0.072、0.089。【结论】半监督RF模型的标准差最小(0.004)且平均AUC值(0.971)最高,展现出最优稳定性与模型预测能力,表明半监督采样方法对模型的优化效果最好。研究成果可为滑坡易发性预测中滑坡负样本选择和模型构建提供参考,同时也为岷江上游流域滑坡风险与防灾减灾提供理论支撑。

关键词: 滑坡易发性预测; 采样策略; 随机森林; 半监督法; 模型平均法; 岷江流域; 滑坡; 影响因素

DOI: 10.13928/j.cnki.wrahe.2025.09.004

开放科学(资源服务)标志码(OSID):

中图分类号: P642.22

文献标志码: A

文章编号: 1000-0860(2025)09-0042-18



收稿日期: 2025-02-19; 修回日期: 2025-04-03; 录用日期: 2025-04-04; 网络出版日期: 2025-04-18

基金项目: 西藏自治区科技计划项目(XZ202401ZY0097); 国家自然科学基金项目(41907228); 四川省科技计划项目(2023YFS0364, 2024YFHZ0154)

作者简介: 骆飞(1999—),男,硕士研究生,主要从事地质灾害及其防治研究。E-mail: 1051651171@qq.com

通信作者: 凌斯祥(1988—),男,副教授,博士,主要从事工程地质灾害与环境的教学、科研工作。E-mail: lingsx@swjtu.edu.cn

©Editorial Department of Water Resources and Hydropower Engineering. This is an open access article under the CC BY-NC-ND license.

Large-scale landslide susceptibility evaluation in upper reaches of Minjiang River based on different selection methods for landslide negative samples and random forest algorithm

LUO Fei¹, LING Sixiang^{1,2}, GAO Fengxin³, LIN Zuhao¹, SUN Chunwei⁴, GAO Fangfang⁵, WU Xiyong^{1,2}

(1. Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, Sichuan, China; 2. Sichuan Province Engineering Technology Research Center of Ecological Mitigation of Geohazards in Tibet Plateau Transportation Corridors, Southwest Jiaotong University, Chengdu 611756, Sichuan, China; 3. College of Engineering, Tibet University, Lhasa 850000, Xizang, China; 4. School of Emergency Management, Xihua University, Chengdu 610065, Sichuan, China; 5. Chengdu Jincheng University, Chengdu 611731, Sichuan, China)

Abstract: Abstract: [Objective] Accurate landslide susceptibility results enable precise prevention and control of landslide hazards and risks. In landslide susceptibility evaluation, selection methods for different landslide negative samples represent a critical uncertainty factor that affects the prediction accuracy of landslide susceptibility. [Methods] Taking the mountainous area of the Minjiang River basin in Sichuan Province as the study area, data on 881 large landslides ($>106 \text{ m}^3$) were compiled through remote sensing imagery. Thirteen landslide evaluation factors including topography and geomorphology, basic geology, hydrogeology, geological environment, seismic parameters, and human activities were selected, and factor redundancy was examined through collinearity analysis. Subsequently, landslide negative samples were selected using random sampling across the study area, sampling in slope zones below 10° , sampling in areas outside 1 km buffer zones around landslides, Information Value (IV) method, Support Vector Machine (SVM) method, and semi-supervised method, with the same proportion as landslide positive samples. These negative samples were further coupled with the Random Forest (RF) to establish Random RF, Low-Slope RF, Buffer RF, IV-RF, SVM-RF, and Semi-Supervised RF models for landslide susceptibility zoning. Finally, the prediction accuracy of different sampling models for landslide negative samples was compared and evaluated using the mean Area Under the Curve (AUC) value derived from the Receiver Operating Characteristic (ROC) curve. [Results] The results showed that: (1) the high and extremely high landslide susceptibility zones obtained by different sampling methods for landslide negative samples were predominantly concentrated on both sides of the river valleys from Wolong to Yingxiu, Miansi to Gu'ergou, Heihu to Musu, and Diexi to Songpan. (2) The prediction accuracy of landslide susceptibility using different sampling methods for landslide negative samples ranked as follows: Semi-Supervised RF ($\overline{AUC}=0.971$) $>$ SVM-RF ($\overline{AUC}=0.954$) $>$ IV-RF ($\overline{AUC}=0.945$) $>$ Buffer RF ($\overline{AUC}=0.902$) $>$ Low-Slope RF ($\overline{AUC}=0.895$) $>$ Random RF ($\overline{AUC}=0.882$). (3) The selection of landslide negative samples in low-susceptibility areas significantly enhanced the prediction accuracy of susceptibility. Compared to Random RF model, the values of the Low-Slope RF, Buffer RF, IV-RF, SVM-RF, and Semi-Supervised RF models increased by 0.013, 0.02, 0.063, 0.072, and 0.089, respectively. [Conclusion] The Semi-Supervised RF model exhibits the smallest standard deviation (0.004) and the highest mean AUC value (0.971), demonstrating optimal stability and prediction capability. This indicates that the semi-supervised sampling method offers the best optimization for the model. These research findings provide references for selecting landslide negative samples and establishing models in landslide susceptibility prediction, while offering theoretical support for landslide risk assessment and disaster mitigation strategies in the upper Minjiang River Basin.

Keywords: landslide susceptibility prediction; sampling strategy; random forest; semi-supervised method; model averaging method; Minjiang River Basin; landslides; influencing factors

0 引言

滑坡灾害具有分布范围广、数量多、危害严重等特点,对人类生命财产安全具有严重威胁^[1]。滑坡易发性预测是滑坡风险评估与潜在滑坡定位的重要基础工作。随着 GIS 技术与机器学习方法的快速发展,滑坡易发性预测向着多学科交叉结合的方式发展^[2-3]。我国山区的基础设施建设对滑坡易发性有重

要影响,如山区交通廊道建设、库区岸坡加固以及城镇扩张等已成为保障人民生命安全的重要因素。

目前,滑坡易发性预测的建模工作包括以下几个关键步骤^[4]: (1)收集目标研究区域的历史滑坡数据及相应的环境因子;(2)确定滑坡正负样本数据集;(3)确定滑坡易发性预测模型;(4)讨论滑坡易发性分区结果。滑坡易发性的建模流程中存在各种不确定性因素,如滑坡样本、滑坡负样本、样本形态、环境

因子与驱动模型方法等, 其过程的不确定性将导致滑坡易发性结果的差异性。在样本形态和驱动模型方法上, 黄发明等^[5] 讨论了以滑坡点、缓冲圆、解译滑坡多边形的不同滑坡样本形态对滑坡易发性预测结果的影响; 李文彬等^[6] 探讨了不同联接方法和数据驱动模型的滑坡易发性预测结果的不确定性; 张圻凯等^[7]、MERGHADI 等^[8] 阐释了单一模型与耦合模型的滑坡易发性预测结果差异性。基于环境因子选择问题, 黄发明等^[9] 讨论了环境因子划分的属性区间数量对滑坡易发性预测结果的影响; 崔玉龙等^[10] 提出了环境因子优化目标空间外向化采样法, 解决多因子作为约束条件取样中容易引入与滑坡相关性差的环境因子的不确定性。现今对于滑坡样本的选择主要依据历史滑坡、遥感影像与现场验证方式进行选取, 其存在的不确定性较小^[11], 如吕蓓茹等^[12] 对比了不同比例的滑坡正负样本集对滑坡易发性的预测结果, 揭示了最优比例的正负样本数据集。对于非滑坡样本(滑坡负样本)的选择引入的不确定性较大, 以往采用全区随机选取的滑坡负样本^[7-13], 该方法无法确保其负样本的代表性。研究滑坡负样本选取策略的成果较少, 如 CHOI 等^[14] 提出了在河道、沟谷等低坡度区域选择滑坡负样本的方法, 其方法可提供负样本的稳定性, 但易导致在低坡度区域过度采样; 缪亚敏等^[15] 提出了缓冲区控制采样, 但缓冲区距离亦可导致滑坡易发性预测结果的不确定性增加; 郭衍昊等^[16]、周晓婷等^[17] 和黄发明等^[18] 提出了不同方法选择负样本与不同模型方法结合的策略, 对比研究了不同负样本方法对滑坡易发性结果的不确定性。虽然上述成果推进了滑坡负样本选择研究, 但多使用单一统计方法对滑坡负样本进行抽样分析, 未详细开展不同方法的滑坡负样本选取策略的对比研究, 仍需进一步全面分析不同滑坡负样本选取策略对滑坡易发性建模结果的影响。

本文选取岷江上游流域为研究区, 采用全区随机采样、低坡度($< 10^\circ$)采样和缓冲区($> 1 \text{ km}$)外采样的滑坡负样本选取策略与不需要负样本的信息量(IV)、仅需少量标记样本即可提升学习能力的半监督法以及监督的支持向量机(SVM)的极低和低易发区进行滑坡负样本选取, 进一步与随机森林(RF)耦合分别构建随机 RF、低坡度 RF、缓冲区 RF、IV-RF、半监督 RF 和 SVM-RF 开展滑坡易发性预测, 并采用 ROC 曲线结合模型平均法对比以上六种模型预测结果的精确度, 探讨不同滑坡负样本选取策略对滑坡易发性预测结果的影响。本研究可为滑坡易发性评价建模过程中的滑坡负样本不确定性提供理论依据,

亦可为滑坡风险评估提供基础。

1 研究区概况及数据源

1.1 岷江流域地质概况

研究区位于我国四川省中北部的岷江上游流域, 地处川西北高原和成都平原过渡的地形陡变带, 形成了高山峡谷地貌(见图 1)。岷江上游流域的高程位于 770~5 931 m 之间, 地势呈现出西北高(若尔盖高原)和东南低(四川盆地)的特点^[19]。岷江上游流域的降雨主要集中在 6—9 月, 其中松潘县至汶川段的多年平均年降水量为 695~1 067 mm。研究区内出露的地层岩性主要为千枚岩、片岩、板岩、灰岩和火成岩, 且发育多条区域型大断裂, 包括映秀—北川断裂、茂县—汶川断裂、松坪沟断裂、岷江断裂等。研究区遭受 1933 年叠溪地震、2008 年汶川地震和 2017 年九寨沟地震的侵袭破坏^[20-22], 且受人类交通工程活动与库区水利水电活动的持续影响, 研究区的滑坡灾害频发。

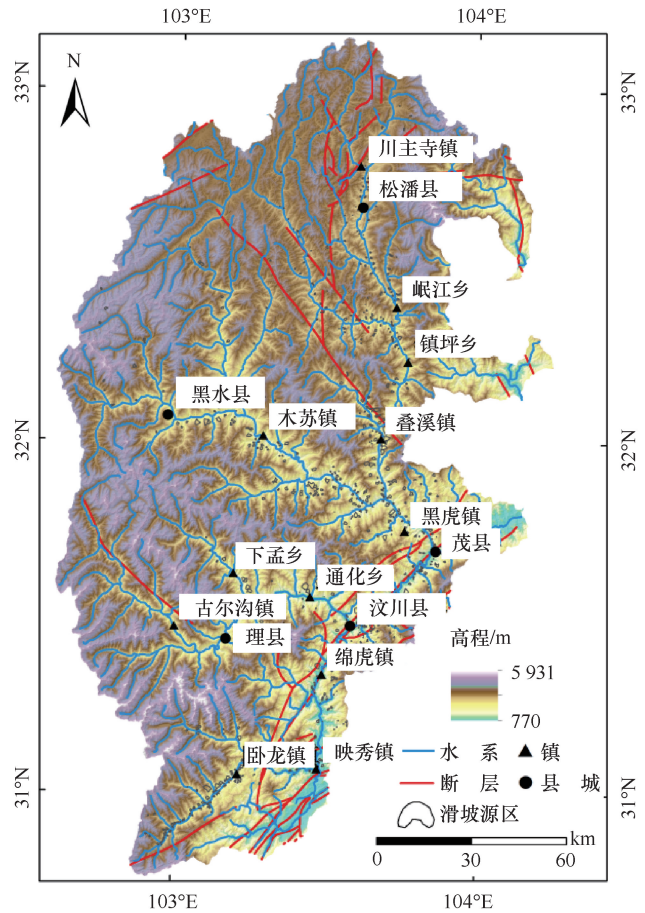


图 1 研究区地质概况及滑坡编录

Fig. 1 Geographical overview and landslide inventory in the upper reaches of Minjiang River

1.2 数据源

本文基于岷江上游流域历史滑坡资料与多期遥感影像数据,在卧龙至映秀段、棉虎至古尔沟段、黑虎至木苏段、叠溪至松潘段遥感解译共 3 861 个滑坡,构建岷江上游流域滑坡灾害编录数据库,从中选取了 881 个大型滑坡($>10^6 \text{ m}^3$)样本进入滑坡数据集(见图 1),其中野外核对了 668 个。本文以滑坡的滑源区作为滑坡样本进行易发性预测,采用 30 m×30 m 栅格单元进行滑坡易发性评价,可有效反映地形地貌特征,符合国家基础空间数据库建设标准,避免栅格单元过多导致计算困难^[7]。因此,对 881 个大型滑坡的滑源区多边形面数据进行 GIS 处理,按 30 像元划分为共计 24 067 个滑坡栅格单元作为滑坡样本进行训练与验证滑坡易发性预测,各项滑坡易发性预测的数据来源如表 1 所列。

2 研究方法

2.1 滑坡负样本选取策略

本文主要采用六种不同的滑坡负样本的选取策略,对比研究不同滑坡负样本选取策略对滑坡易发性结果预测的不确定性(见图 2)。滑坡负样本生成的数量与滑坡样本数量相同,两者均为 24 067 个栅格单元,其滑坡负样本选取策略的方式如下。

(1)全区随机采样:在岷江上游流域的全区域随机生成滑坡负样本[见图 2(a)]。

(2)低坡度采样:依据岷江上游流域坡度特征与滑坡易发性结果综合判断,设置坡度 $< 10^\circ$ 为坡度阈值,选择坡度 $< 10^\circ$ 的低坡度区域随机生成滑坡负样本[见图 2(b)]。

(3)缓冲区采样:以识别的滑坡边界作为缓冲区的起点,在滑坡多边形边界以外 1 km 的缓冲区随机生成滑坡负样本[见图 2(c)]。

(4)信息量(IV)采样:利用物理统计的叠加原理

计算出各滑坡评价因子中不同子类的信息量值,以自然间断点法划分极低、低、中、高和极高滑坡易发性分区,再从极低与低滑坡易发性区域随机生成滑坡负样本[见图 2(d)]。

(5)支持向量机(SVM)采样:将滑坡正样本与全区随机采样的滑坡负样本作为数据集输入 SVM 模型,计算滑坡易发性预测结果,以自然间断点法划分极低、低、中、高和极高滑坡易发性分区,再从极低与低的滑坡易发性区域随机选取滑坡负样本[见图 2(e)]。

(6)半监督采样:随机森林属于监督式学习算法,通过融合自训练、协同训练、图结构学习等技术,可扩展为半监督算法^[23-24]。本次融合自训练技术实现随机森林从监督式算法到半监督式算法的转化,即先用历史滑坡数据训练初始随机森林模型,预测滑坡极低与低易发性区域,从中筛选高置信度的非滑坡样本,实现半监督采样,将高置信度的非滑坡样本加入训练集并重新训练模型,实现迭代优化[见图 2(f)]。

2.2 随机森林模型

随机森林(Random Forest, RF)是一种集成学习方法(见图 3),采用多棵独立训练的决策树(Decision Tree)进行预测,每棵树都是弱学习器,随机森林通过组合多个弱学习器,可降低过拟合现象,且具有优良的准确性和稳定性,对大规模数据集和高维特征空间的适应能力强,其异常值和噪声的鲁棒性较好,并且相对于单棵决策树,不需要过多的参数调优工作^[25-26]。随机森林方法主要步骤如下。

(1)自助采样(Bootstrap Sampling):每棵决策树可使用自助采样,生成训练数据集,但可能导致每棵树的训练集重叠和差异。

(2)随机特征选择:每棵树在训练过程中随机选择部分特征进行建模,将有助于减少建模模型的方法

表 1 数据来源

Table 1 Data sources

数 据	来 源
高程	美国地质调查局(https://www.usgs.gov/)
坡度、坡向、曲率、地形起伏度、地表粗糙度、地表切割度、高程变异系数、地形湿度指数	基于 DEM 数据(30 m 分辨率)
断层、地层岩性	1:20 万区域地质图(http://www.igeodata.org)
道路、河流、土地利用类型	全国地理信息资源目录服务(http://www.globallandcover.com/)
多年年均降雨量	国家气象信息中心(https://data.cma.cn/)
地震峰值加速度	中国地震动参数区划图(https://gb18306.net/oldmap)
归一化植被指数(NDVI)	美国国家航空航天局(https://search.earthdata.nasa.gov/)
滑坡边界与滑源区边界数据	历史滑坡编录及遥感解译(https://www.gscloud.cn/)

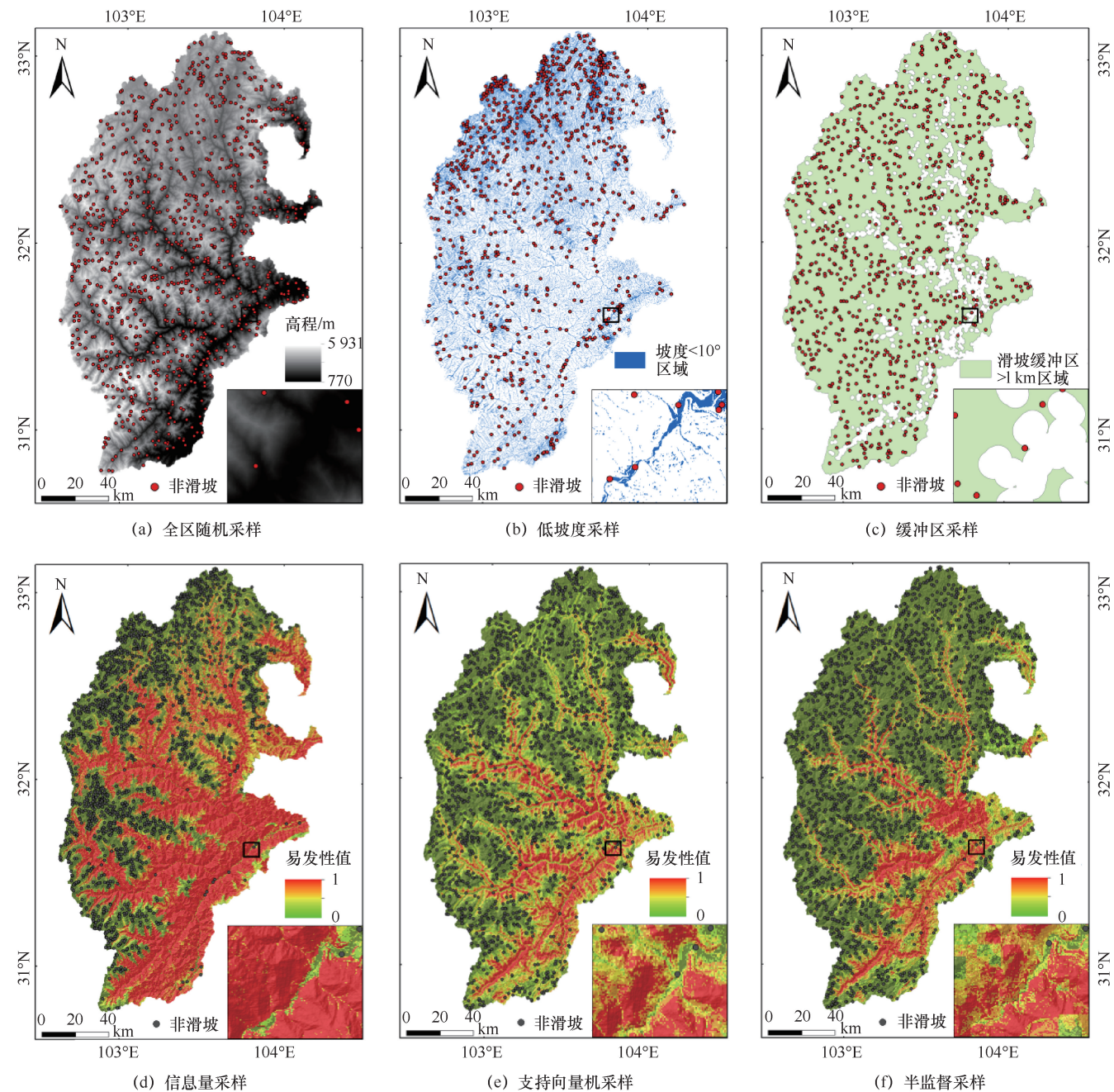


图 2 滑坡负样本选取策略示意

Fig. 2 Schematic diagram of different selection methods for landslide negative samples

差, 提升泛化能力。

(3) 决策树集成: 随机森林由多棵决策树组成, 每棵树独立训练, 互不相关和互不干扰。

(4) 投票机制: 对于分类问题, 多棵树投票决定最终分类结果; 而对于回归问题, 则取所有树的平均预测值作为最终输出结果。

2.3 滑坡易发性建模精度评价

ROC 曲线 (Receiver Operating Characteristic Curve, 简称 ROC) 广泛应用于滑坡易发性的精度评

价, 其横坐标为特异性 (FPR), 纵坐标为敏感性 (TPR), 可展示模型在不同阈值下的分类效果^[26-29]。ROC 曲线下的面积 (Area Under Curve, 简称 AUC) 可量化模型性能的分类精度, 当 AUC 值越接近 1, 说明模型的易发性预测性能越好, 计算公式为^[4, 26-29]

$$\begin{cases} TPR = \frac{TP}{TP + FN} \\ FPR = \frac{FP}{FP + TN} \end{cases} \quad (1)$$

式中, TP 为正确分类的滑坡样本数量; FN 为错误分类的非滑坡样本数量; FP 为错误分类的滑坡样本数量; TN 为正确分类的非滑坡样本数量。

以往研究中, 多关注 ROC 曲线的任意一次或最好一次的 AUC 值来评价模型的易发性精度^[29], 而忽略了模型易发性建模预测的不确定性, 可能导致遗漏部分极高和高滑坡易发性分区。为更好对比评价本文的不同滑坡负样本选取策略对滑坡易发性预测的影响, 降低滑坡易发性预测的不确定性, 将各滑坡负样本选取策略的 AUC 值进行多次平均, 采用 AUC 均值 (\overline{AUC}) 代表模型评价精度, 标准差代表模型的稳定性。本文采用模型平均法将易发性概率矩阵加权平均得到平均化后的概率矩阵^[30], 再利用 ROC 曲线量化评价精度, 其计算公式为

$$\begin{cases} \overline{Y} = \frac{1}{n} \sum_{i=1}^n w_i Y_i \\ \sigma = \sqrt{\frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n}} \end{cases} \quad (2)$$

式中, n 为随机评价次数, 本文取 10; w_i 为第 i 个随机评价结果的权重, 本文假设各评价结果权重相同, 均为 1; Y_i 为第 i 个随机评价的易发性值; \overline{Y} 为 n 次随机评价取均值后的易发性值, 即 \overline{AUC} 值; σ 为标准差值。

2.4 滑坡易发性评价流程

本研究旨在对比六种不同滑坡负样本选取策略对滑坡易发性结果的不确定性, 其评价流程如图 4 所示。

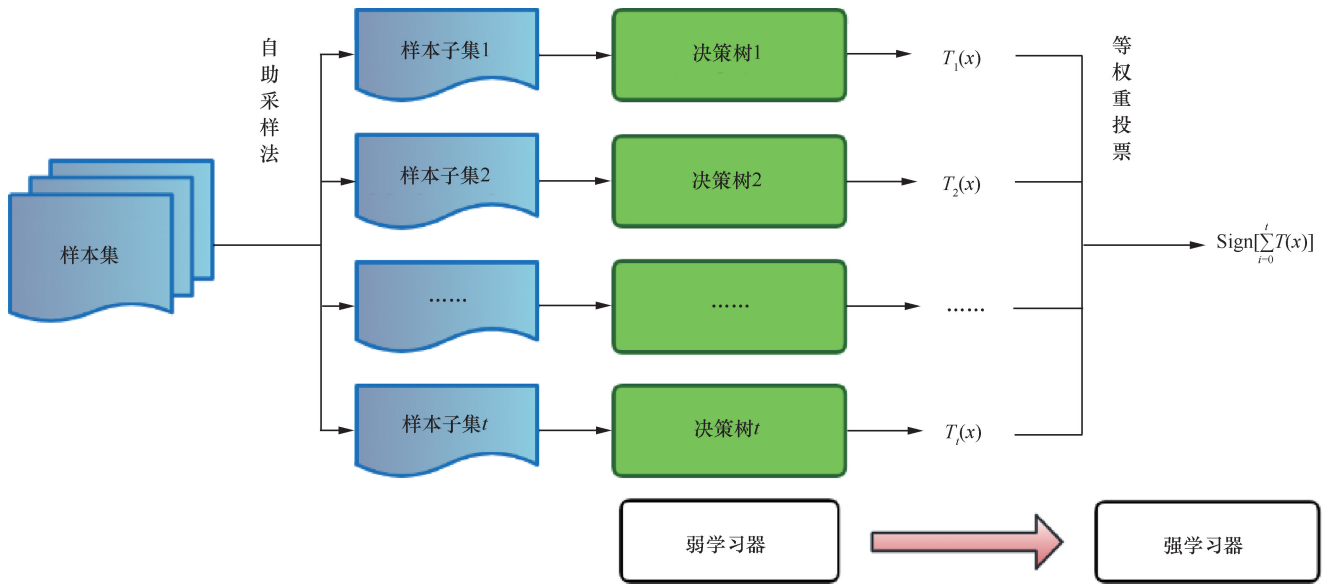


图 3 随机森林原理示意

Fig. 3 Schematic diagram of random forest principle

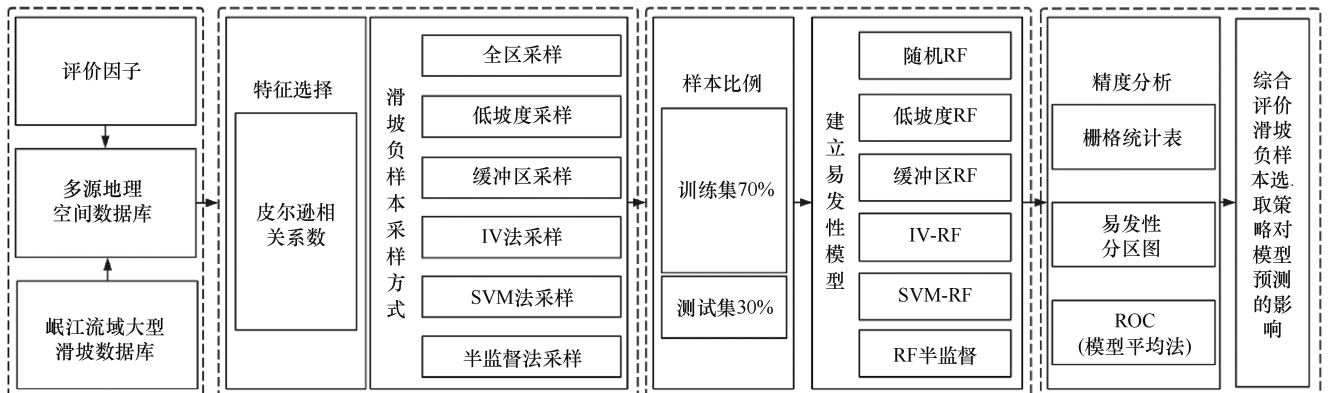


图 4 不同滑坡负样本选择方式的滑坡易发性预测建模流程

Fig. 4 Modeling workflow for landslide susceptibility prediction under different selection methods for landslide negative samples

(1)收集岷江上游流域滑坡数据、滑坡遥感识别与野外调查验证,建立滑坡正样本数据集,选取滑坡易发性评价因子数据进行相关性分析,确定滑坡易发性评价因子与滑坡之间的关系。

(2)采取全区随机采样、低坡度采样、缓冲区采样、IV 采样、SVM 采样与半监督法采样的策略生产与滑坡正样本相同数量的滑坡负样本。

(3)将六种不同滑坡负样本选取策略产生的滑坡负样本与滑坡正样本进行合并,分别按 7:3 随机划分为滑坡易发性训练集和测试集,再将 RF 模型连接形成全区随机 RF、低坡度 RF、缓冲区 RF、IV-RF、SVM-RF 与半监督 RF 模型开展滑坡易发性预测。

(4)对不同滑坡负样本选取策略情景下的滑坡易发性结果进行分区,利用栅格统计表、ROC 曲线的 AUC 平均值(\overline{AUC})及因子重要性分析不同滑坡负样本选取策略下的滑坡易发性建模的不确定性。

3 岷江上游滑坡指标因子分析

3.1 评价因子相关性分析

为保证评价因子的客观性和统计继承性,从数据源中提取了地形地貌、基础地质、水文地质、地质环境与诱发条件因素方面共 17 个评价因子数据(见图 5)。地形地貌选取了 8 个评价因子,分别为海拔(ELEV)、坡度(SLP)、坡向(ASP)、曲率(CURV)、地形起伏度(TRI)、地表粗糙度(SR)、地表切割度(SI)和高程变异系数(ECV);基础地质因子为距断层距离(DTF)和地层岩性(LITH);水文地质因子选取了地形湿度指数(TWI)和距河流距离(DTH);地质环境因子由植被覆盖指数(NDVI)和土地利用类型(LUT)组成;外界触发因子为地震峰值加速度(PGA)、年均降雨量(MAP)和距道路距离(HEA)。

当选择滑坡易发性评价因子时,必须确保所选评价因子间相互独立,不能出现因子多重共线性^[4]。为保证评价因子间的相关性,采用皮尔逊相关系数研

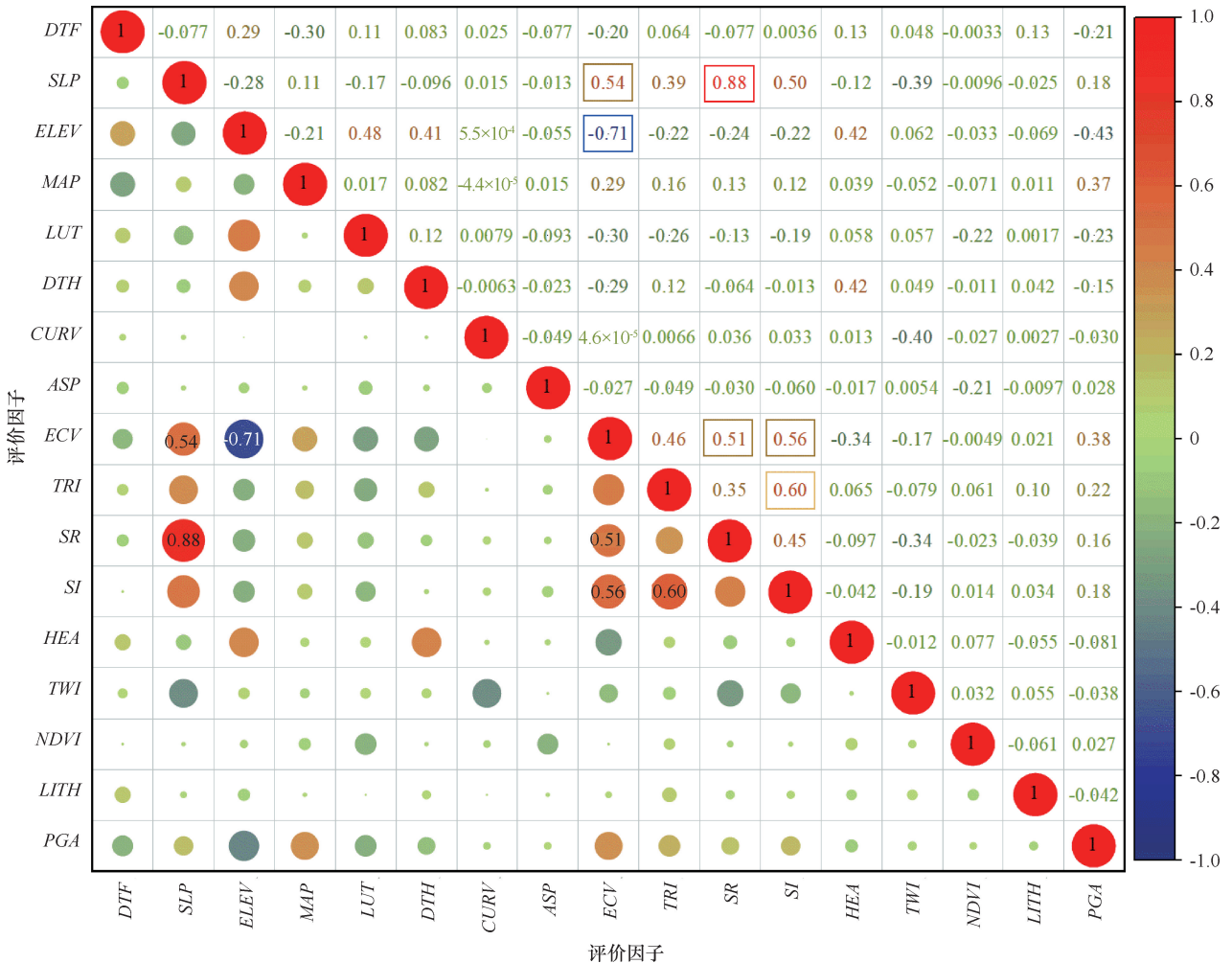


图 5 皮尔逊相关系数热力图

Fig. 5 Heatmap of Pearson correlation coefficients

究各因子间的相关性(见图5),当皮尔逊相关系数绝对值大于0.5时,表明两因子间存在显著相关性^[4],应从评价因子体系中移除。高程变异系数与高程、坡度、地表切割度、地表粗糙度的相关系数分别为-0.71、0.54、0.56、0.51,地面粗糙度与坡度为0.88,地表切割度与地形起伏度为0.60,均大于0.5的临界值,表明存在强相关性(见图5)。高程变异系数、地表粗糙度和地表切割度作为地形数据衍生数据,且95.2%的滑坡发生于林地和草地(占研究区93.7%的面积),表明土地利用类型无法全面反映其对滑坡发生概率的影响,不符合科学性和全面性原则,故将高程变异系数、地表粗糙度、地表切割度和土地利用类型从评价因子中剔除。最终,选取高程、坡度、坡向、曲率、地形起伏度、距断层距离、地层岩性、距河流距离、植被覆盖指数、地震峰值加速度、地形湿度指数、年均降雨量和距道路距离共13个评价因子作为滑坡易发性建模的特征因子(见表2)。

3.2 因子频率比分析

频率比(FR)是基于数理统计的预测方法,其比值是由选取评价因子分类区间的滑坡比除以面积比计算得到,可定量反映各评价因子与滑坡易发性间的关系^[7,31],其计算公式为

$$FR = \frac{n/N}{m/M} \quad (3)$$

式中, n 为因子分类区间滑坡发生的面积; N 为总滑坡面积; m 为因子分类区间的面积; M 为研究区的总面积。

当 $FR > 1$ 时,表示该因子所在区间有利于滑坡发生;而当 $FR < 1$ 时,则表明该区间不利于滑坡发生;当 $FR = 1$ 时,表明结果难以判断^[4,32]。经过相关性检验后的评价因子如图6所示,其对应 FR 值如表2所列。

(1)地形地貌因素。地形地貌因素均从30 m DEM中获取,各地形地貌因素分类如图6(a)—(e)所示,计算获取的频率比值如表2所列。岷江上游流域高程因子如图6(a)所示,在1370~3170 m范围内的频率比值(FR)大于1,而在770~1370 m和3170~5931 m时的 FR 值小于1,整体而言,滑坡数量随着海拔的升高具有先升高而后降低的趋势,在中山地貌至高山地貌海拔过渡区(1370~3170 m)发生滑坡概率较大。坡度因子如图6(b)所示,坡度在大于30°的 FR 值大于1,而坡度小于30°时的 FR 值小于1, FR 值随着坡度增加而增大,表明坡度越陡滑坡发生的概率越大。坡向因子如图6(c)所示,当坡向为西至西南向时, FR

值大于1,在东至北向时, FR 值小于1,表明滑坡易发生在西至西南的向阳坡(见表2),主要是由于向阳坡能接受更多日照时长与日照强度,促使斜坡更易风化破碎^[7]。曲率在1~59和-53~-2时, FR 值大于1,曲率在-2~1时, FR 值小于1(见表2),说明边坡表面凹凸不平程度越大,滑坡发生的概率越高[见图6(d)]。地形起伏度反映区域地形的起伏变化[见图6(e)],地形起伏度的 FR 随着地形起伏度增大而增大,地形起伏度在750~1812 m时, FR 值大于1,在0~750 m时, FR 值小于1(见表2),表明在高起伏度地区发生滑坡的概率更大。

(2)基础地质因素。断层破碎带是滑坡易发地带,以距断层的距离作为基础地质因素的量化指标之一,如图6(f)所示。距断层距离在0~2500 m时, FR 值大于1,且距断层越近时, FR 值越大,在距断层距离>2500 m时, FR 值小于1,指示离断层越近的区域越易发生滑坡(见表2)。地层岩性决定了岩石的力学特性、结构特征、水软化作用和地震作用的反应^[29],地层岩性因子如图6(g)所示,当地层为志留系、泥盆系、石炭系与二叠系的岩土体, FR 值大于1,指示该地层发生滑坡的概率更大(见表2)。

(3)水文地质因素。地形湿度指数(TWI)指示了流域内岩土体水分的湿润程度,当 TWI 在1.8~3.8与5.8~7.8时, FR 值大于1,在3.8~5.8与大于7.8时, FR 值小于1(见表2),表明在中等土壤湿度区域易发生滑坡。河流水系对斜坡岩土体长期冲刷、侧蚀和浸润作用,将增加边坡临空面,降低斜坡稳定性,因此,以距河流的距离作为水文地质因素的量化指标之一^[30],距河流的距离因子如图6(h)所示。距河流距离在0~1600 m时, FR 值大于1,且距河流越近时的 FR 值越大,在大于1600 m时, FR 小于1,表明离河流越近的区域易发生滑坡(见表2)。

(4)地表覆盖因素。植被覆盖指数($NDVI$)指示了研究区的植被生长状况和覆盖范围^[33-34],其 $NDVI$ 因子如图6(i)所示,当 $NDVI$ 值在-3.95~0.3时, FR 值大于1,在0.3~0.76时, FR 值小于1(见表2),说明在植被越少的区域更易发生滑坡。

(5)诱发条件因素。地震可破坏岩土体的结构和斜坡稳定性,诱发滑坡灾害,以地震动峰值加速度(PGA)来衡量地震活动强度[见图6(g)],当 PGA 大于0.20g时, FR 值大于1,在0.04g~0.20g时, FR 值小于1,指示在地震活跃区的滑坡发生概率较高(见表2)。降雨将促使岩土体软化与增加渗透压力,诱发斜坡失稳滑动^[34],年均降雨量可作为滑坡易

表2 研究区滑坡易发性评价因子指标分级

Table 2 Classification of evaluation factors of landslide susceptibility in the study area

影响因子	变量值	类型	栅格比例/%	滑坡数/个	FR值	影响因子	变量值	类型	栅格比例/%	滑坡数/个	FR值
高程/m	770~1 370	连续型	1.07	9	0.95	曲率	-53~-8	连续型	0.43	9	2.37
	1 370~1 970		4.51	143	3.60		-8~-4		3.97	47	1.34
	1 970~2 570		9.52	254	3.03		-4~-2		11.46	120	1.19
	2 570~3 170		17.14	303	2.01		-2~-1		14.08	110	0.89
	3 170~3 770		32.39	147	0.52		-1~1		42.35	327	0.88
	3 770~4 370		27.53	24	0.10		1~2		12.79	119	1.06
	4 370~4 970		7.42	1	0.02		2~4		10.70	101	1.07
	4 970~5 931		0.42	0	0.00		4~8		3.78	39	1.17
坡度/(°)	0~10	连续型	5.16	3	0.07	地形起伏度/m	8~59.333	连续型	0.44	9	2.32
	10~20		16.14	47	0.33		0~250		2.03	2	0.11
	20~30		28.75	186	0.73		250~500		15.17	61	0.46
	30~40		29.85	320	1.22		500~750		33.51	246	0.83
	40~50		15.56	233	1.70		750~900		21.63	218	1.14
	50~60		4.04	81	2.27		900~1 150		22.13	275	1.41
	>60		0.50	11	2.49		1 150~1 400		5.13	70	1.55
									1 400~1 812	0.39	9
坡向/(°)	平地	离散型	0.16	1	0.70	断层距离/m	<500	连续型	4.48	69	1.75
	北		11.26	86	0.87		500~1 000		4.29	63	1.67
	东北		12.04	96	0.90		1 000~1 500		3.96	56	1.60
	东		14.09	127	0.92		1 500~2 000		3.66	42	1.30
	东南		13.38	100	0.85		2 000~2 500		3.37	49	1.65
	南		11.98	103	0.98		2 500~3 000		3.09	24	0.88
	西南		11.87	130	1.24		3 000~3 500		2.87	23	0.91
	西北		12.47	107	0.97		3 500~4 000		2.73	36	0.95
TWI	1.8~3.8	连续型	8.76	133	1.72	水系距离/m	>4 000	连续型	71.55	519	0.82
	3.8~5.8		56.52	446	0.90		0~400		13.97	143	1.16
	5.8~7.8		23.64	244	1.17		400~800		13.58	286	2.39
	7.8~9.8		6.90	51	0.84		800~1 200		12.98	159	1.39
	9.8~11.8		2.14	6	0.32		1 200~1 600		12.08	107	1.01
11.8<	2.05	1	0.06	1 600~2 000	10.85	57	0.60				
NDVI	-3.95~0	连续型	51.73	475	1.04	年均降雨量/mm	>2 000	连续型	36.54	129	0.40
	0~0.1		20.51	146	1.11		695~730		5.83	75	0.92
	0.1~0.2		10.45	108	1.17		730~760		22.39	251	0.89
	0.2~0.3		6.71	71	1.20		760~790		25.01	147	0.67
	0.3~0.4		4.92	32	0.74		790~820		23.53	209	1.01
	0.4~0.5		3.43	29	0.96		820~850		16.54	123	0.84
	0.5~0.6		1.82	11	0.69		850~880		4.35	37	1.27
	0.6~0.76		0.45	9	0.87		880~1 067		2.35	39	1.88
地层岩性	Q	离散型	0.79	9	0.74	道路距离/m	0~300	连续型	4.66	77	1.88
	N		0.05	3	0.44		300~600		4.41	148	3.81
	E		0.23	0	0		600~900		4.26	133	3.55
	J		2.46	0	0		900~1 200		4.11	86	2.38
	T ₃		46.92	216	0.52		1 200~1 500		3.97	62	1.77
	T ₂		14.40	78	0.61	>1 500	78.60	375	0.54		
	T ₁		4.55	31	0.77	PGA/g	0.04~0.2	连续型	69.13	348	0.57
	P		9.59	167	1.98		0.2~0.4		14.34	173	1.37
	C		3.31	32	1.10		0.4~0.6		5.50	70	1.44
	D		8.31	144	1.97		0.6~0.8		3.16	96	3.45
	S		7.18	133	2.10		0.8~1		3.48	125	4.08
	O		0.03	0	0		1~1.2		2.81	54	2.18
	ε		0.11	0	0		1.2~1.72		1.59	15	1.07
	Z		2.02	68	0.83						

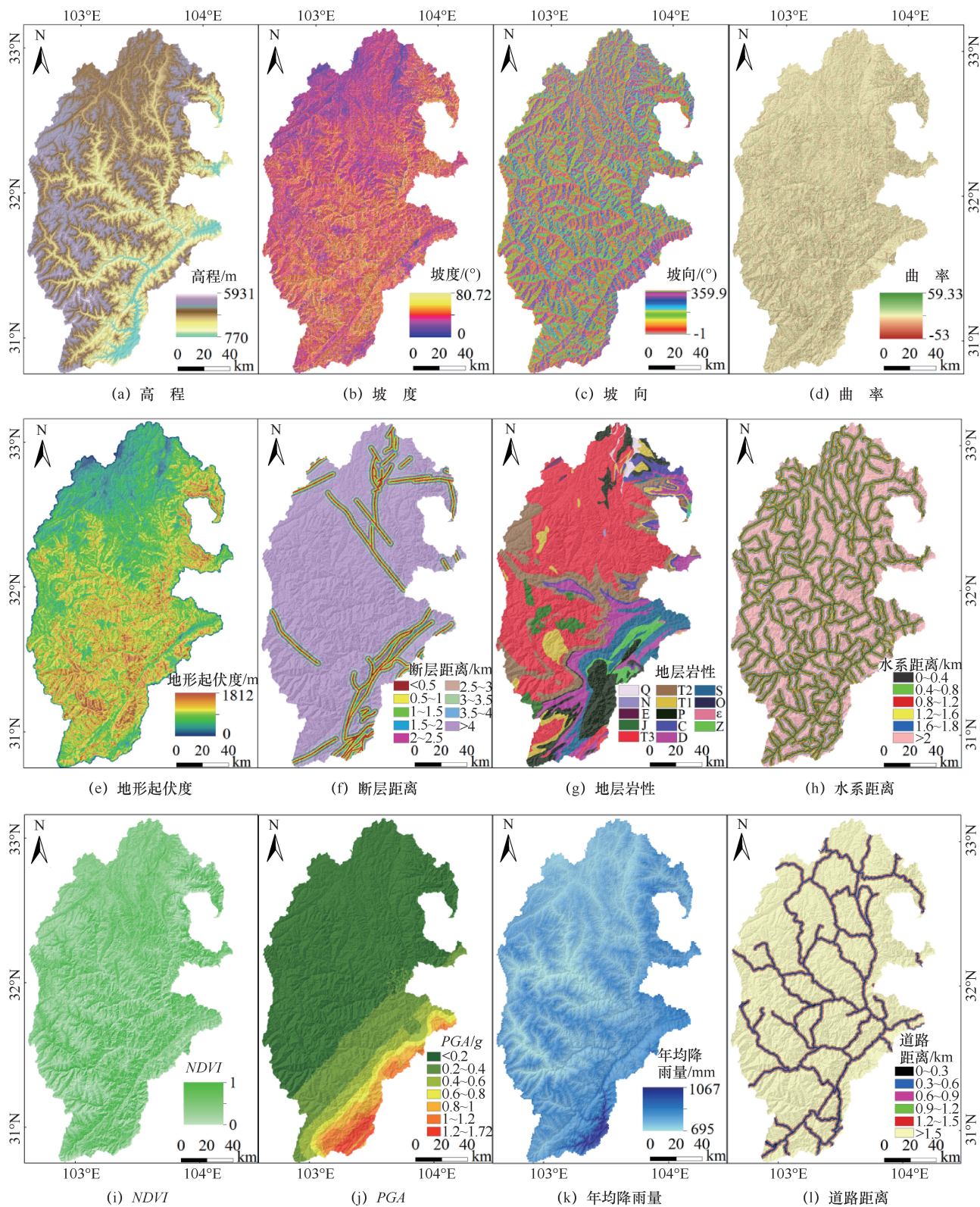


图 6 滑坡易发性评价因子

Fig. 6 Evaluation factors of landslide susceptibility

发性评价因子[见图 6(k)], 当年降雨量在 790~820 mm 与 850~1 067 mm 时, FR 值大于 1, 而在 695~790 mm 与 820~850 mm 时, FR 值小于 1(见表 2)。

人类工程活动也是滑坡活动的诱发因素之一, 本文以距道路距离衡量人类工程活动[见图 6(l)], 当距道路距离为 0~1500 m 时, FR 值大于 1, 且距道路距离

表 3 不同滑坡负样本选择策略的滑坡易发性等级统计结果

Table 3 Statistical results of landslide susceptibility grades under different selection methods for landslide negative samples

指 标	易发性级别	随机 RF	低坡度 RF	缓冲区 RF	IV-RF	SVM-RF	半监督 RF
易发性分区面积 百分比/%	极 低	61.58	60.69	61.59	37.80	49.03	59.06
	低	10.15	11.34	11.17	5.28	13.07	7.52
	中	8.49	8.10	8.01	5.72	7.75	5.34
	高	7.76	7.84	7.63	11.35	9.23	6.36
	极 高	12.02	12.03	11.61	39.85	20.92	21.72
滑坡面积 百分比/%	极 低	2.12	3.19	2.80	5.02	2.63	0.46
	低	4.76	4.15	4.33	1.60	2.79	0.84
	中	5.07	4.49	4.50	1.84	3.82	1.30
	高	10.97	11.39	11.87	5.05	11.21	3.40
	极 高	77.08	76.78	76.50	86.50	79.54	94.01

越近, FR 值越大, 而当距道路距离大于 1 500 m 时, FR 值小于 1, 说明距道路越近生滑坡的概率越大(见表 2)。

4 岷江上游滑坡易发性评价

本文通过随机 RF、低坡度 RF、缓冲区 RF、IV-RF、SVM-RF 和半监督 RF 模型计算岷江上游滑坡的极高与高易发区合计面积分别占整个研究区面积的 19.78%、19.87%、19.23%、51.19%、28.08%、23.08%(见表 3 和图 7)。不同滑坡负样本采样策略耦合 RF 模型获取的极高和高滑坡易发区主要集中于以下四处位置(见图 8): (1)卧龙镇至映秀镇段, 主要位于汶川地震发震断层(北川—映秀断裂)上盘, PGA 值在 $0.80 g \sim 1.72 g$, 处于高地震动峰值加速度区域, 且距离震中距离均在 14 km 以内, 是汶川地震灾区中距离震中最近、震害最严重的区域之一^[19]; (2)绵虬镇—通化乡—古尔沟镇段, 该段位于杂谷脑河段的高山峡谷地带, 地层岩性主要为千枚岩、板岩等, 在降雨、地震以及河流下切侵蚀作用下有利于增加滑坡发生概率; (3)黑虎镇至木苏镇段, 该段山高谷深, 沟壑纵横, 大部分地区地形起伏度在 1 150~1 812 m, 坡度大于 40° , 地形起伏大且坡度较陡, 地层岩性主要由千枚岩、板岩等组成, 易受地震活动导致岩层破碎, 形成滑坡高风险地区; (4)叠溪镇至松潘段, 该段地处岷江断裂带, 地层岩性主要为千枚岩、板岩等, 叠溪镇松坪沟乡曾发生里氏 7.5 级地震^[35], 导致岩层破碎, 诱发大型滑坡, 且该区高程在 1 370~4 390 m 之间, 海拔高低悬殊大, 形成滑坡高风险地区。

滑坡面积在随机 RF、低坡度 RF、缓冲区 RF、IV-RF、SVM-RF 和半监督 RF 模型中预测出的极高和高易发分区的滑坡面积所占的比例分别为 88.05%、88.17%、88.38%、91.55%、90.75%、97.41%, 表明本文的不同滑坡负样本采样策略耦合的 RF 模型模

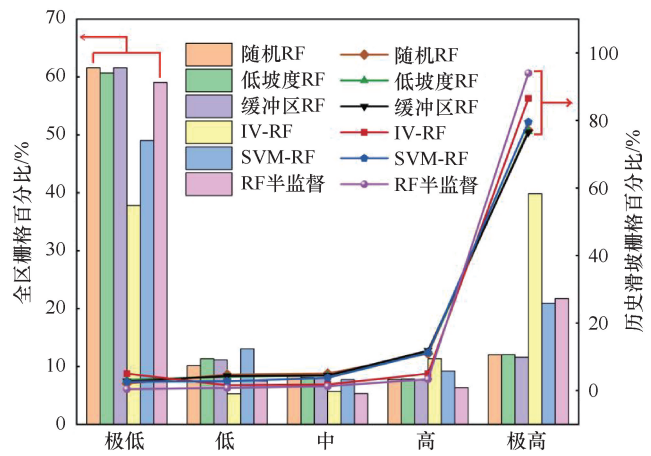


图 7 六种模型栅格统计

Fig. 7 Grid statistics of six models

型预测的极高和高滑坡易发性区域均覆盖了大多数滑坡目录清单。随机 RF、低坡度 RF 与缓冲区 RF 模型预测的滑坡易发性分区在各滑坡面积占比情况相似(见图 8), 表明低坡度的滑坡负样本采样与缓冲区滑坡负样本采样的优化效果不明显。IV-RF 模型预测的极高和高滑坡易发性的滑坡面积占比达 91.55%, 而极高和高滑坡易发性分区面积占比达 51.19%, 与实际情况不符(见图 7)。同时, IV-RF 模型将小金河、牟托村以及牟托沟沟谷等不发生滑坡的区域也划分为高和极高易发区, 模型可解释性差[见图 8(d)], 主要是 IV 的滑坡负样本采样时多集中于山脊处, 导致提取滑坡负样本质量降低, 可能是 IV 的滑坡负样本采样忽视了指标间的相互作用与联系而影响滑坡易发性预测结果的全面性和深度^[4]。SVM-RF 模型相较传统的随机 RF 模型在极高和高滑坡易发性分区面积占比提高划分了约 10%, 且滑坡面积占比提高了约 2.7%, 表明 SVM 的滑坡负样本采样对滑坡易发性预测结果具有一定的优化效果(见图 7)。半监督 RF 模型相较传统的随机 RF 模型在极高和高滑坡易发性

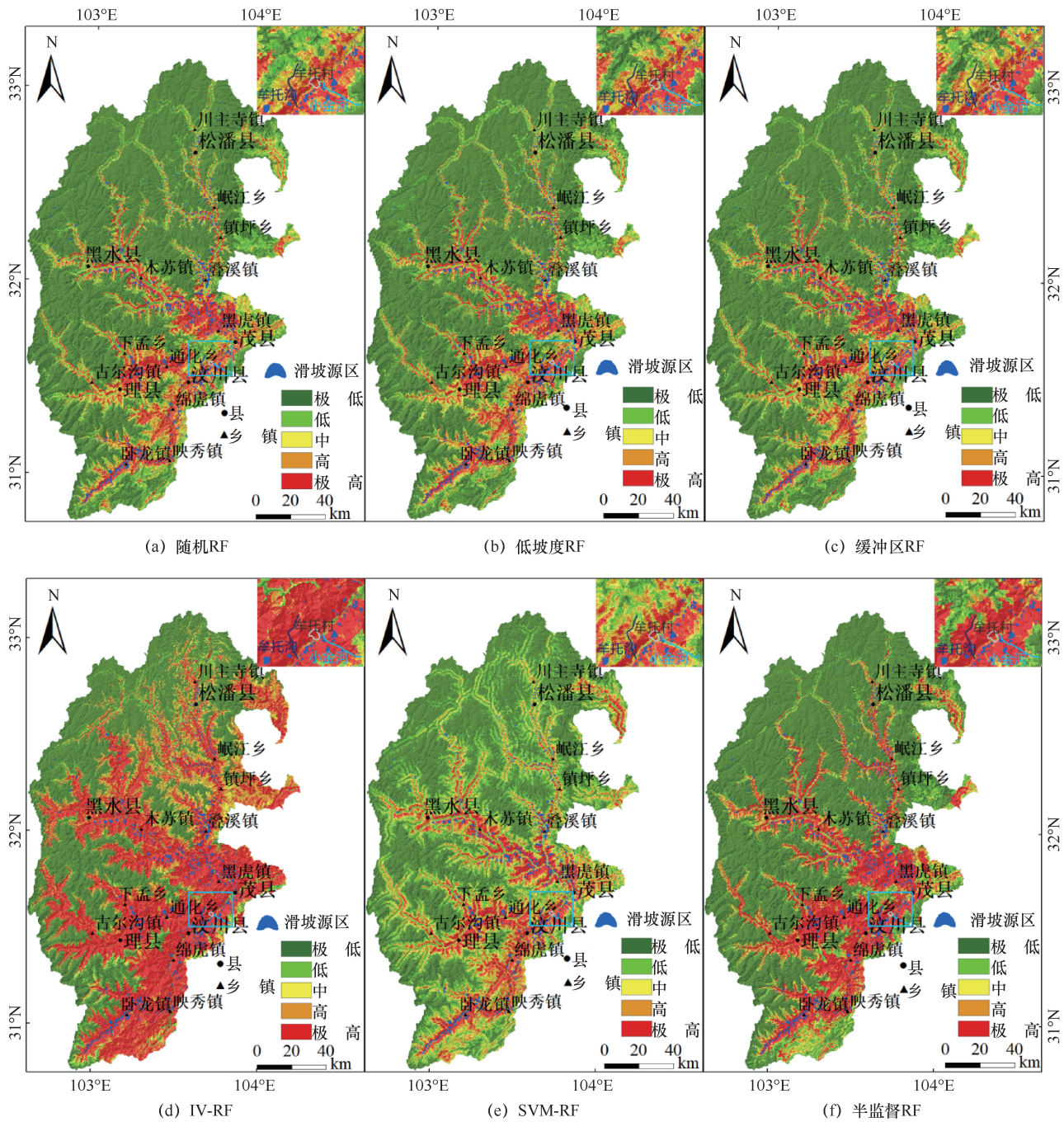


图8 基于不同负样本的预测模型易发性分区

Fig. 8 Susceptibility zoning of prediction models based on different negative samples

分区面积占比提高划分了约8%，而滑坡面积占比提高了约9%，达97.41%，远超其余五种模型，初步表明半监督的滑坡负样本采样对滑坡易发性预测结果的优化效果最好(见图7)。

5 讨论与分析

5.1 不同滑坡负样本选取策略的滑坡易发性精度评价

随机RF、低坡度RF、缓冲区RF、IV-RF、

SVM-RF与半监督RF模型的滑坡易发性预测AUC值范围分别为0.868~0.896、0.881~0.911、0.874~0.916、0.936~0.952、0.941~0.965、0.963~0.978(见图9)，而加权平均的 \overline{AUC} 值分别为0.882、0.895、0.902、0.945、0.954、0.971，表明六种不同滑坡负样本取样策略的滑坡易发性模型均有较好预测精度。随机RF模型的 \overline{AUC} 值为0.882，标准差为0.009，上下限AUC值相差0.028，最小AUC值为0.868，表明随机RF模型具有良好的预测精度与稳

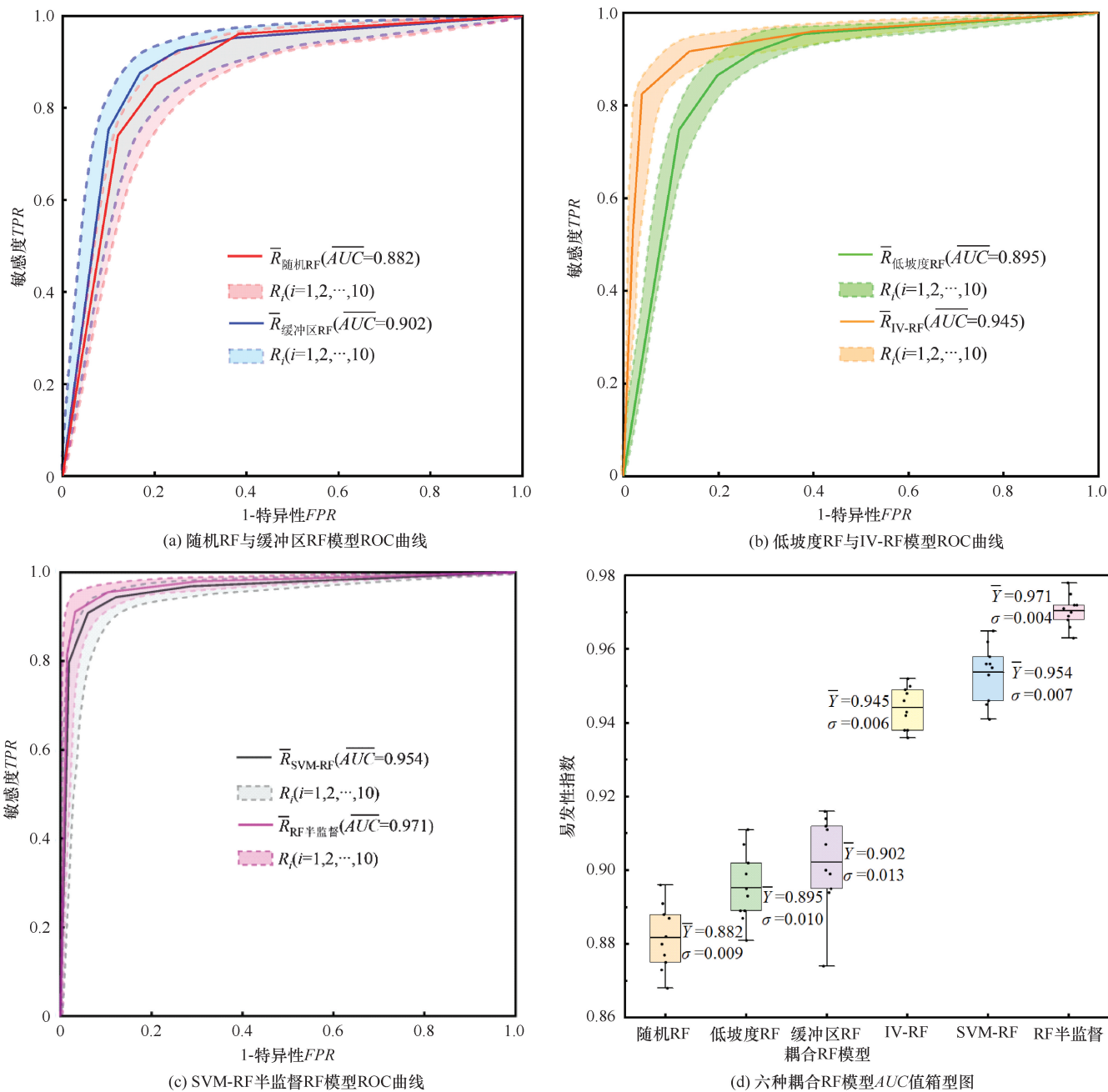


图9 各模型 ROC 曲线与 AUC 值箱型图

Fig. 9 Box plots of ROC curves and AUC values of different models

定性。低坡度 RF 相较于随机 RF 在精度上有轻微提升, \overline{AUC} 值提高了 0.013, 达到 0.895, 但 AUC 值箱型图分布较为分散, 低坡度 RF 模型的标准差 ($\sigma = 0.010$) 略大于随机 RF, 稳定性稍差, 表明在岷江流域选用低坡度的滑坡负样本采样法对滑坡易发性模型优化效果有限, 主要是研究区均为大型滑坡, 在坡度 $< 10^\circ$ 区域内也有部分滑坡分布, 导致选取的滑坡负样本质量不高, 影响模型预测结果(见图 9)。缓冲区 RF 模型与低坡度 RF 模型的加权 \overline{AUC} 值相近, 相较

于随机 RF 的加权 \overline{AUC} 值提高了 0.020, 但缓冲区 RF 模型的标准差 σ 为 0.013, AUC 值上下限相差 0.042, 具有最大的上下限跨度, 其稳定性最差, 说明在滑坡缓冲区 1 km 外的区域选择滑坡负样本对模型的有一定优化, 但将会降低模型的稳定性, 主要是缓冲区范围目前尚未有统一的滑坡负样本选取标准^[17-18], 导致所选取的滑坡负样本质量起伏较大, 而影响滑坡易发性预测结果。

基于统计方法的 IV-RF 模型相较于随机 RF、低坡度 RF 与缓冲区 RF 模型的加权 \overline{AUC} 值具有明显的

提升, 其 \overline{AUC} 值达到 0.945, 标准差 σ 为 0.006, 表明 IV-RF 模型具有较高的稳定性, 但从易发性分区面积百分比与易发性分区可知 IV-RF 模型预测结果可解释性差, 与岷江流域滑坡分布具有一定矛盾[见图 8(d)]。基于耦合机器学习的 SVM-RF 模型相较于随机 RF、低坡度 RF、缓冲区 RF 和 IV-RF 模型, 其精度与模型稳定性都有较大的提升, 加权 \overline{AUC} 值达到 0.954, 标准差 σ 为 0.007, 其模型的稳定性良好, 同时 SVM-RF 模型的极高与高易发区覆盖了研究区大部分滑坡目录清单, 模型结果可解释性良好, SVM 的滑坡负样本采样方法对模型的优化效果显著。半监督 RF 模型相较于以上五种模型, 加权 \overline{AUC} 值的精度提升最大, \overline{AUC} 值达到 0.971, 标准差为 0.004, 显示出最优精度与最高稳定性, 结合易发性分区图, 表明半监督 RF 模型具有优秀的滑坡易发性预测结果, 说明半监督的滑坡负样本采样方法对滑坡易发性建模的优化效果最好[见图 9(c)—(d)]。

通过对不同滑坡负样本选择策略的易发性等级统计结果、易发性分区、预测精度以及预测结果稳定性的对比分析, 随机选取滑坡负样本的 RF 模型表现尚可, 且因其操作简单, 预测精度即可满足部分工程情况, 多数研究目前依旧采取随机采样^[7,22,27], 但该方法所选样本质量低, 可能包含高坡度、近滑坡区等不合理的负样本, 导致模型混淆滑坡与非滑坡的边界特征, 预测精度依赖于当次所选负样本的质量, 不确定性高。低坡度选取滑坡负样本的 RF 模型精度有一定提升, 但其稳定性下降, 说明低坡度区域滑坡发生概率低的认识缺乏普适性依据, 可能忽略低坡度区域的潜在滑坡, 如泥石流触发区, 以及该方法受限于研究区, 在陡峭山区低坡度区域面积有限, 导致负样本代表性不足。基于滑坡 1 km 缓冲区外选取负样本的 RF 模型, 精度有一定提升但稳定性最差, 该方法虽排除滑坡邻近区域, 避免正负样本空间重叠, 但目前滑坡缓冲区最佳阈值设定尚无统一标准, 不同区域需动态调整^[15], 模型泛化能力较差。利用信息量量化因子与滑坡的关系, 在低易发区选取滑坡负样本的 IV-RF 模型, 预测精度与稳定性都有一定程度提高, 但该方法依赖于信息量法的准确分区, 在复杂山区, 信息量法难以辨别各相关因子之间的相关性^[4]与联系, 而导致过多的划分与实际不符的滑坡高易发区, 虽精度有所提升, 但模型可解释性差。通过 SVM 初步预测分区, 在低易发区选取滑坡负样本的 SVM-RF 模型, 预测精度与稳定性都有了

极大提升, 因其 SVM 法能适应因子间复杂关系, 捕捉非线性因子交互作用, 初步预测分区更加准确, 但其计算复杂, 需额外训练 SVM 模型, 增加计算时间和资源消耗, 且该方法依赖于初始标签, 若初始训练集存在偏差, 可能影响负样本选择。通过半监督法选取滑坡负样本的半监督 RF 模型, 极高易发区覆盖 94.01% 的历史滑坡, 预测精度与模型稳定性都显著优于其他方法, 因为半监督法通过迭代优化筛选高置信度负样本, 减少对初始标签的依赖, 且随机森林算法对异常值和噪声的鲁棒性较好, 无须过多优化^[23-25], 工作量也相对较小, 但该方法对初始模型更加敏感, 若初始模型预测误差较大, 迭代过程可能放大偏差。六种不同滑坡负样本采样策略的预测精度都表现出良好的效果, 但其适用场景与局限性有所不同, 随机采样、低坡度采样、缓冲区采样适合快速部署, IV 法采样、SVM 法采样、半监督法采样适用于精细化评估, 未来可结合多源数据, 如 InSAR 形变监测、优化负样本选择, 探索自适应阈值设定, 提升模型跨区域迁移能力。

5.2 评价因子重要性分析

评价因子的重要性分析有助于分析研究区极高与高易发性区域主要受控因素^[29-33], 加深对研究区的滑坡灾害认识。高程因子在随机 RF、低坡度 RF、缓冲区 RF、IV-RF、SVM-RF 与半监督 RF 模型中均为主控因子, 重要性权重占比分别为 19.4%、18.9%、19.9%、19.0%、42.2% 和 33.9% (见图 10)。高程在 1 370~3 170 m 范围内覆盖了 79.45% 的滑坡目录 (见表 2), 说明高程是岷江流域滑坡发生最为重要的特征因子。随机 RF、低坡度 RF、缓冲区 RF 和 IV-RF 模型的评价因子重要性柱状图相似, 距断层距离、坡向、年均降雨量、地形起伏度、地层岩性与 PGA 在模型中较为重要, 重要性权重占比在 7.7%~9.2% 之间, 其余因子的重要性权重占比相对较小, 在 3.5%~6.2% 之间[见图 10(a)—(d)], 但低坡度 RF 模型的坡度因子重要性权重有显著提高, 达到 9.0%, 主要原因是根据坡度特性确定阈值来选择滑坡负样本。SVM-RF 模型高程重要性权重占比最大, 达到 42.2%, 地形起伏度次之 (19.2%), 其余特征因子重要性权重占比相对偏小, 表明 SVM-RF 模型对地形地貌因素信息更为敏感。半监督 RF 模型除高程占比 33.9% 以外, PGA 重要性权重占比可达 12.6%, 断层距离、年均降雨量、地形起伏度与道路距离因子较为重要, 重要性权重占比在 6.3%~8.4% 之间, 其余因子重要性相对较小, 表明半监督 RF 模型受地形

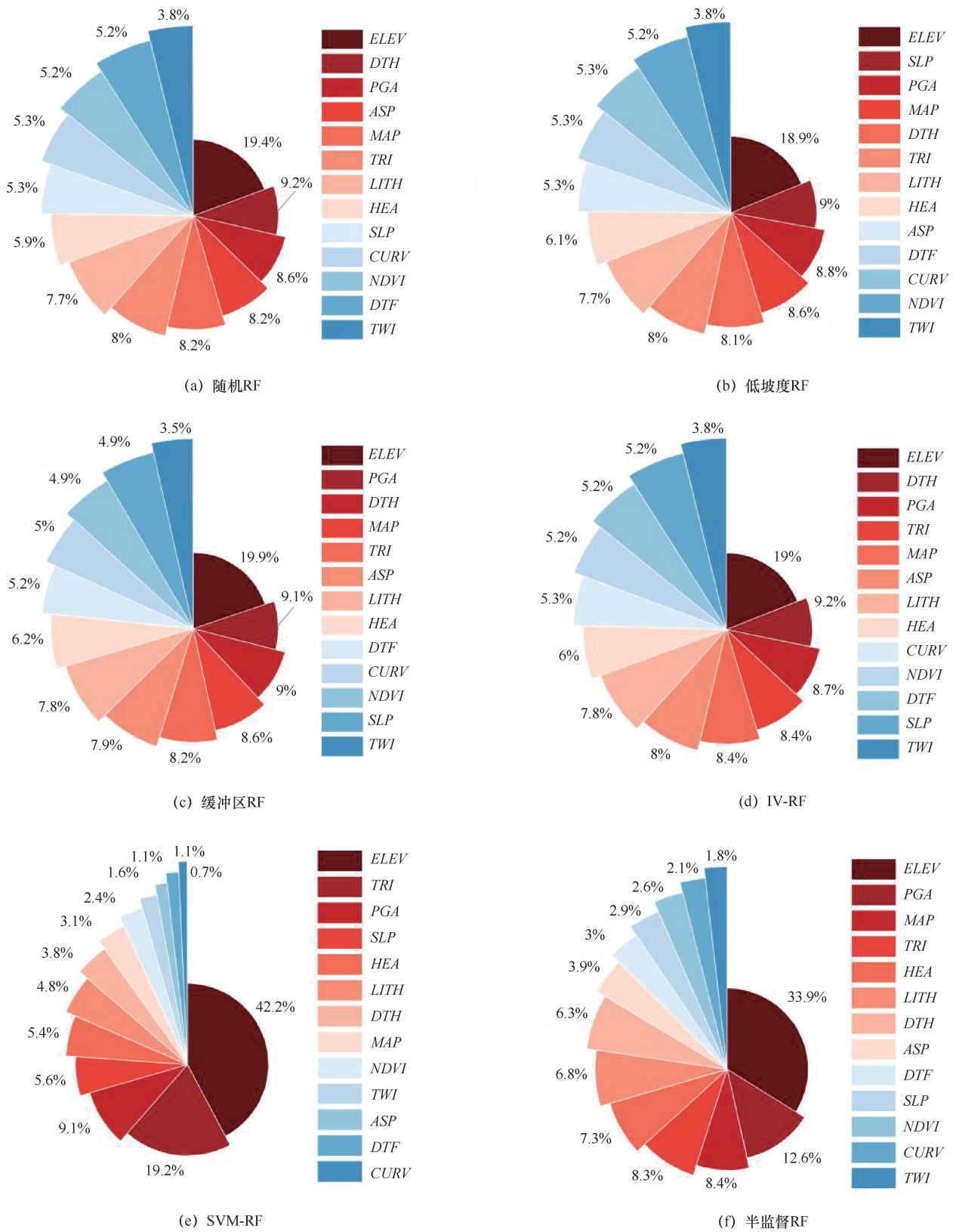


图 10 六种模型环境因子重要性

Fig. 10 Importance of environmental factors in six models

地貌与诱发条件因素综合影响更为敏感。在 3.2 节频率比分析的评价因子指标分级表中, 地形地貌因素包

括高程、坡度、曲率、地形起伏度评价因子及诱发条件因素包括 *PGA*、道路距离评价因子的 *FR* 值均有达

到 2~3 的区间段, 而 FR 值大于 1 说明该区段有利于滑坡发生, FR 值越大该区段滑坡发生概率越大^[7], 表明岷江上游流域滑坡对地形地貌与诱发条件因素更敏感, 结合滑坡易发性预测模型结果的精度与稳定性, 半监督 RF 模型均表现出最优结果。综合考虑因子频率比分析与滑坡易发性预测结果, 半监督 RF 模型的因子重要性排序更加科学, 在模型预测时所占的权重更符合实际。

6 结论

(1) 随机 RF、低坡度 RF、缓冲区 RF、IV-RF、SVM-RF 与半监督 RF 模型所预测出的极高和高易发性分区主要分布于卧龙至映秀段、绵虬至古尔沟段、黑虎至木苏段以及叠溪至川主寺段, 主要属于高山峡谷地带, 其年均降雨量较大, 受断裂与地震活动的影响较大, 易演变为滑坡灾害。

(2) 与全区域随机采样方法相比, 低坡度采样、缓冲区采样、IV 法采样、SVM 法采样与半监督法采样法在模型预测精度和稳定性方面均表现出不同程度的优化效果。各模型的加权 \overline{AUC} 值分别为随机 RF (0.882) < 低坡度 RF (0.895) < 缓冲区 RF (0.902) < IV-RF (0.945) < SVM-RF (0.954) < 半监督 RF (0.971), 且标准差表现为半监督 RF (0.004) < IV-RF (0.006) < SVM-RF (0.007) < 随机 RF (0.009) < 低坡度 RF (0.010) < 缓冲区 RF (0.013), 说明不同滑坡负样本选取策略可优化滑坡负样本的质量, 有利于提高滑坡易发性建模性能, 且在选取最终滑坡负样本数据集前, 先对研究区进行初次预测分区, 从低与极低易发区选择滑坡负样本可明显提升滑坡易发性建模精度, 模型的稳定性和泛化性更优秀。

(3) 基于机器学习采样法的 SVM-RF 模型与半监督 RF 模型表现突出, 加权 \overline{AUC} 值达到 0.954 与 0.971, 标准差 σ 仅为 0.006 与 0.004, 模型在保证高精度的同时有良好的稳定性, 且评价因子重要性占比更符合岷江流域地理特征, 可用较少的极高与高易发区覆盖更多的滑坡目录。基于机器学习的 SVM 采样法与半监督采样法优于全区随机采样法、低坡度采样法、缓冲区采样法以及基于统计方法的 IV 采样法, 其中半监督的滑坡负样本采样法对滑坡易发性建模表现出最好和最稳定的优化效果。

参考文献 (References):

[1] 崔鹏, 贾洋, 苏风环, 等. 青藏高原自然灾害发育现状与未来关注的科学问题[J]. 中国科学院院刊, 2017, 32(9): 985-992.

CUI Peng, JIA Yang, SUN Fenghuan, et al. Natural hazards in Tibetan Plateau and key issue for feature research [J]. Bulletin of Chinese Academy of Sciences, 2017, 32(9): 985-992.

[2] MIN D H, KIM Y, KIM S, et al. Strategy of oversampling geotechnical parameters through geostatistical, SMOTE, and CTGAN methods for assessing susceptibility of landslide [J]. Landslides, 2024, 21(2): 291-307.

[3] JI Jian, CUI Hongzhi, ZHANG Tong, et al. A GIS-based tool for probabilistic physical modelling and prediction of landslides; GIS-FORM landslide susceptibility analysis in seismic areas [J]. Landslides, 2022, 19(9): 2213-2231.

[4] HUANG Junpeng, WU Xiyong, LING Sixiang, et al. A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020 [J]. Environmental Science and Pollution Research, 2022, 29(58): 86954-86993.

[5] 黄发明, 曹昱, 范宣梅, 等. 不同滑坡边界及其空间形状对滑坡易发性预测不确定性的影响规律[J]. 岩石力学与工程学报, 2021, 40(S2): 3227-3240.

HUANG Faming, CAO Yu, FAN Xuanmei, et al. Effects of different landslide boundaries and their spatial shapes on the uncertainty of landslide susceptibility prediction [J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(S2): 3227-3240.

[6] 李文彬, 范宣梅, 黄发明, 等. 不同环境因子联接和预测模型的滑坡易发性建模不确定性[J]. 地球科学, 2021, 46(10): 3777-3795.

LI Wenbin, FAN Xuanmei, HUANG Faming, et al. Modeling uncertainty of landslide susceptibility in connection and prediction models of different environmental factors [J]. Earth Science, 2021, 46(10): 3777-3795.

[7] 张纪恺, 凌斯祥, 李晓宁, 等. 九寨沟县滑坡灾害易发性快速评估模型对比研究[J]. 岩石力学与工程学报, 2020, 39(8): 1595-1610.

ZHANG Qikai, LING Sixiang, LI Xiaoning, et al. Comparison of rapid assessment models for landslide hazard susceptibility in Jiuzhaigou County [J]. Chinese Journal of Rock Mechanics and Engineering, 2020, 39(8): 1595-1610.

[8] 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 Review, 2020, 207: 103225.

[9] 黄发明, 叶舟, 姚池, 等. 滑坡易发性预测不确定性: 环境因子不同属性区间划分和不同数据驱动模型的影响[J]. 地球科学, 2020, 45(12): 4535-4549.

HUANG Faming, YE Zhou, YAO Chi, et al. Uncertainty in landslide susceptibility prediction: the effect of different attribute intervals of environmental factors and different data-driven models [J]. Earth Science, 2020, 45(12): 4535-4549.

[10] 崔玉龙, 朱路路, 徐敏, 等. 基于环境因子优化 TSES 法选择负样本及其在滑坡易发性评价中的应用[J]. 地质科技通报, 2024, 43(3): 192-199.

CUI Yulong, ZHU Lulu, XU Min, et al. Selection of negative

- samples based on environmental factors optimization TSES method and its application in landslide susceptibility evaluation[J]. *Bulletin of Geological Science and Technology*, 2024, 43(3): 192-199.
- [11] ZHAO P X, MASOUMI Z, KALANTARI M, et al. A GIS-based landslide susceptibility mapping and variable importance analysis using artificial intelligent training-based methods [J]. *Remote Sensing*, 2022, 14(1): 211.
- [12] 吕蓓茹, 彭玲, 李樵民. 顾及样本敏感性的滑坡易发性评价[J]. *测绘通报*, 2022(11): 20-25.
- LYU Beiru, PENG Ling, LI Qiaomin. Landslide susceptibility evaluation considering sample sensitivity[J]. *Bulletin of Surveying and Mapping*, 2022(11): 20-25.
- [13] BADAVATH N, SAHOO S, SAMAL R. Landslide susceptibility mapping for West-Jaintia Hills district, Meghalaya[J]. *Sadh ana- Proceedings of the School of Engineering Sciences*, 2024, 49(1): 1-19.
- [14] CHOI J, OH H J, LEE H J, et al. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS[J]. *Engineering Geology*, 2012, 124(1): 12-23.
- [15] 缪亚敏, 朱阿兴, 杨琳, 等. 滑坡危险度评价对 BCS 负样本采样的敏感性[J]. *山地学报*, 2016, 34(4): 432-441.
- MIAO Yamin, ZHU Axing, YANG Lin, et al. Sensitivity of landslide risk assessment to BCS negative sample sampling[J]. *Journal of Mountain Science*, 2016, 34(4): 432-441.
- [16] 郭衍昊, 窦杰, 向子林, 等. 基于优化负样本采样策略的梯度提升决策树与随机森林的汶川同震滑坡易发性评价[J]. *地质科技通报*, 2024, 43(3): 251-265.
- GUO Yanhao, DOU Jie, XIANG Zilin, et al. Evaluation of susceptibility of Wenchuan coearthquake landslide using gradient lifting decision tree and random forest based on optimal negative sample sampling strategy[J]. *Bulletin of Geological Science and Technology*, 2024, 43(3): 251-265.
- [17] 周晓亭, 黄发明, 吴伟成, 等. 基于耦合信息量法选择负样本的区域滑坡易发性预测[J]. *工程科学与技术*, 2022, 54(3): 25-35.
- ZHOU Xiaoting, HUANG Faming, WU Weicheng, et al. Regional landslide susceptibility prediction based on negative sample selection base on coupled information method [J]. *Engineering Science and Technology*, 2022, 54(3): 25-35.
- [18] 黄发明, 曾诗怡, 姚池, 等. 滑坡易发性预测建模的不确定性: 不同“非滑坡样本”选择方式的影响[J]. *工程科学与技术*, 2024, 56(1): 169-182.
- HUANG Faming, ZENG Shiyi, YAO Chi, et al. Uncertainty in landslide susceptibility prediction modeling: The influence of different “non-landslide sample” selection methods[J]. *Engineering Science and Technology*, 2024, 56(1): 169-182.
- [19] FAN N N, YANG X G, STOROZUM M J, et al. Geoarchaeological evidence of an ancient landslide dam (13~4.7 ka) and consequent outburst flood on the Minjiang River near Wenchuan, China[J]. *Catena*, 2024, 242: 108137.
- [20] QIN Menglin, WU Xinyu, ZHOU Yijia, et al. How is carbon storage in plateau-plain transition zone influenced? Evidence from Minjiang River Basin, China[J]. *Journal of Cleaner Production*, 2025, 491: 144766.
- [21] LIU Xin, ZHONG Bo, LI Naying, et al. Notable ecological risks of microplastics to Minjiang River ecosystem over headwater to upstream in Eastern Qinghai-Tibetan Plateau [J]. *Water Research*, 2025, 274: 123137.
- [22] 李艳豪, 蒋汉朝, 徐红艳, 等. 四川岷江上游滑坡触发因素分析[J]. *地震地质*, 2015, 37(4): 1147-1161.
- LI Yanhao, JIANG Hanchao, XU Hongyan, et al. Analysis of triggering factors of landslides in the upper reaches of the Minjiang River in Sichuan[J]. *Seismology and Geology*, 2015, 37(4): 1147-1161.
- [23] JI Xia, WANG Lingzhu, FANG Xiaohao, et al. Semi-supervised batch active learning based on mutual information [J]. *Applied Intelligence*, 2025, 55(2): 1-18.
- [24] CHEN Hangyu, XIE Xijiong, LI Di, et al. Semi-supervised learning with Deep Laplacian Support Vector Machine[J]. *Pattern Analysis and Applications*, 2025, 28(1): 1-13.
- [25] QI Lihui, WANG Xuedong, CUI Wang, et al. Random forest-based screening of environmental geohazard probability factors in Panshi City, China[J]. *Advances In Space Research*, 2025, 75(1): 414-431.
- [26] 武建. 人工智能技术在水利行业中的应用实践与展望[J]. *水利发展研究*, 2024, 24(8): 44-49.
- WU Jian. Application practice and prospect of artificial intelligence in water sector[J]. *Water Resources Development Research*, 2024, 24(8): 44-49.
- [27] 舒晓燕, 巫锡勇, 文洪, 等. 新疆天山伊阿铁路区域雪崩易发性与潜在释放区识别对比研究[J]. *工程地质学报*, 2023, 31(4): 1200-1212.
- SHU Xiaoyan, WU Xiyong, WEN Hong, et al. A comparative study of avalanche susceptibility and potential release area identification in the Tianshan Mountains, Xinjiang [J]. *Journal of Engineering Geology*, 2019, 31(4): 1200-1212.
- [28] CHICCO D, JURMAN G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification[J]. *Biodata Mining*, 2023, 16(1): 1-23.
- [29] 杨强, 王高峰, 丁伟翠, 等. 多种组合模型的区域滑坡易发性及精度评价[J]. *自然灾害学报*, 2021, 30(2): 36-51.
- YANG Qiang, WANG Gaofeng, DING Weicui, et al. Evaluation of regional landslide susceptibility and accuracy of multiple combined models[J]. *Journal of Natural Disasters*, 2019, 30(2): 36-51.
- [30] 谭林, 张璐璐, 魏鑫, 等. 基于 U-Net 语义分割网络的区域滑坡易发性评价方法和跨地区泛化能力研究[J]. *土木工程学报*, 2024. DOI: 10.15951/j. tmge. 23110923.
- TAN Lin, ZHANG Lulu, WEI Xin, et al. Research on regional landslide susceptibility evaluation method and cross-regional generalization ability based on U-Net semantic segmentation network

[J]. *Journal of Civil Engineering*, 2024. DOI: 10.15951/j.tmgcxb. 23110923.

- [31] HUANG Faming, CAO Zhongshan, GUO Jianfei, et al. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping[J]. *Catena*, 2020, 191: 104580.
- [32] LI Langping, LAN Hengxing, GUO Changbao, et al. A modified frequency ratio method for landslide susceptibility assessment [J]. *Landslides*, 2017, 14(2): 727-741.
- [33] LING Sixiang, ZHAO Siyuan, HUANG Junpeng, et al. Landslide susceptibility assessment using statistical and machine learning techniques: A case study in the upper reaches of the Minjiang River, southwestern China [J]. *Frontiers in Earth Science*, 2022, 10: 986172.
- [34] 刘帅, 王涛, 曹佳文, 等. 基于优化随机森林模型的降雨群发滑

坡易发性评价研究: 以西秦岭极端降雨事件为例[J]. *地质通报*, 2024, 43(6): 958-970.

- LIU Shuai, WANG Tao, CAO Jiawen, et al. Evaluation of landslide vulnerability of rainfall cluster based on optimized random forest model: A case study of extreme rainfall events in the West Qinling Mountains[J]. *Geological Bulletin*, 2024, 43(6): 958-970.
- [35] 刘亚静, 刘红健. 基于信息量-随机森林模型的地震带地质灾害易发性评价: 以松潘-较场地震带为例[J]. *科学技术与工程*, 2024, 24(1): 143-154.
- LIU Yajing, LIU Hongjian. Evaluation of geological hazard susceptibility in seismic zones based on information-random forest model: a case study of Songpan-field seismic zone [J]. *Science Technology and Engineering*, 2019, 24(1): 143-154.

(责任编辑 王海锋)