引文格式：郭广彬,田婷,周忠国,等. 多机器学习算法下沧源县滑坡地质灾害易发性评价[J]. 地理空间信息, 2026, 24(2):64-67. doi:10. 3969/j. issn. 1672-4623.2026.02.013

多机器学习算法下沧源县滑坡地质灾害易发性评价

郭广彬[1]，田　婷[2*]，周忠国[1]，王忠伟[3]，李　涞[1]
（1. 云南省设计院集团勘察院有限公司，云南 昆明 650228；2. 云南建投第一勘察设计有限公司，云南 昆明 650102；
3. 保山施勐高速公路投资开发有限公司，云南 保山 678200）

摘　要：针对在滑坡样本稀缺环境下的滑坡地质灾害易发性评价问题，旨在探索一种准确、稳定的机器学习算法。通过对沧源县的滑坡灾害进行分析，本研究选用了LightGBM、随机森林（RF）和支持向量机（SVM）3种算法，并结合11个关键影响因子构建了滑坡易发性评价模型。研究发现，LightGBM模型在预测滑坡危险性区域方面表现最为突出，与实际滑坡点的一致性极高，尤其在滑坡高发的东部区域，其预测准确性得到了充分验证。通过ROC曲线评估，LightGBM模型的AUC值高达0.836，显著优于RF和SVM模型，证实了其在数据稀缺环境下的强大预测能力。此外，关键影响因子分析中揭示河流侵蚀、断层破裂带和人类工程活动对滑坡灾害易发性具有显著影响。因此，建议未来将LightGBM模型作为滑坡地质灾害易发性评价的首选方案，其分析结果可为滑坡灾害的预防和减灾工作提供技术支撑。其次，应关注河流沿岸地带、断层活动区以及人类工程活动频繁区域的边坡稳定性状况，切实采取有效的防护措施并加强监测工作，以确保天然边坡的稳固与安全，并最大限度地降低潜在的灾害风险。

关键词：GIS；滑坡易发性评价；机器学习；沧源县
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2026）02-0064-04

Landslide Geological Hazard Vulnerability Evaluation in Cangyuan County, Lincang City, Yunnan Province Under Multiple Machine Learning Algorithms

GUO Guangbin[1], TIAN Ting[2], ZHOU Zhongguo[1], WANG Zhongwei[3], LI Su[1]
(1. Yunnan Design Institute Group Co., Ltd., Kunming 650228, China;
2. Yunnan Construction Investment First Survey and Design Co., Ltd., Kunming 650102, China;
3. Baoshan Shimeng Expressway Investment and Development Co., Ltd., Baoshan 678200, China)

Abstract: This study focuses on the issue of landslide susceptibility evaluation in environments with scarce landslide samples, aiming to explore an accurate and stable machine learning algorithm. By analyzing landslides in Cangyuan County, Lincang City, Yunnan Province, we selected LightGBM, random forest (RF) and support vector machine (SVM) three algorithms, and constructed a landslide susceptibility model using 11 key influence factors. The results showed that LightGBM model performed most prominently in predicting high-risk landslide areas, demonstrating a high degree of consistency with actual landslide points, especially in the eastern part of the county where landslides were more frequent. The predictive accuracy of LightGBM model was fully validated in these high-risk areas. Through ROC curve evaluation, LightGBM model achieved an AUC value of 0.836, significantly outperforming the RF and SVM models, which confirmed its robust predictive capability in data scarce environments. Furthermore, the analysis of key influence factors revealed that river erosion, fault rupture zones, and human engineering activities had a significant impact on landslide susceptibility. Therefore, it is recommended that LightGBM model be adopted as the preferred approach for landslide susceptibility evaluation, providing technical support for landslide prevention and mitigation efforts. Additionally, attention should be paid to the stability of slopes along riverbanks, fault activity zones, and areas with frequent human engineering activities. Effective protective measures and enhanced monitoring should be implemented to ensure the stability of natural slopes and minimize potential disaster risks.

Key words: GIS, landslide susceptibility evaluation, machine learning, Cangyuan County

当前，滑坡易发性评价是滑坡危险性评价和风险性评价的重要基础[1-3]。易发性评价旨在通过一定的方法和模型，确定某一区域发生滑坡灾害的空间概率，从而为防灾减灾提供科学依据。现有的易发性评价方法多种多样[4]，包括专家打分法、信息量法、证据权法、确定性系数法、频率比法，以及层次分析法、模糊综合评判法和机器学习法等。在这些方法中，机器学习法因其出色的效果而备受青睐[5]，尤其是LightGBM、SVM（支持向量机）和RF（随机森林）这3种算法，在滑坡易发性评价中显示出卓越的性能[6-8]。

在实际情况中，滑坡灾害数据常常不足，尤其是在那些滑坡事件记录有限且缺乏有效数据管理的地

第一作者简介：郭广彬（1986—），本科，从事岩土工程勘察、地质灾害防治、地基基础施工及管理工作，E-mail：18213580520@163.com。
通信作者：田婷（1986—），本科，高级工程师，从事岩土工程勘察、岩土设计及管理工作，E-mail：mermaidsky77@163.com。

区。人工收集和标记滑坡编目数据不仅成本高昂，而且难度极大。这种数据稀缺的情况可能导致传统易发性评价方法因样本不足而产生不准确和不稳定的评估结果。鉴于此，本文深入分析LightGBM、SVM和RF这3种算法在滑坡样本稀缺情况下的表现，以期找到一种在数据稀疏环境中依然能够准确、稳定地进行滑坡易发性评估的机器学习算法，这将为滑坡灾害的防灾减灾工作提供新的思路和解决方案。

1　研究区概况

沧源佤族自治县位于中国西南边陲的云南省西部，具有“边疆、山区、民族、贫困”多重特征的自治县。其地理位置独特，具体位于23°04′~23°40′N、98°52′~99°43′E，北接耿马傣族佤族自治县，东邻双江拉祜族佤族布朗族傣族自治县，东南与澜沧拉祜族自治县相连，西南则与缅甸接壤[9]，拥有长达290.79 km的边界线，研究区概况图见图1。

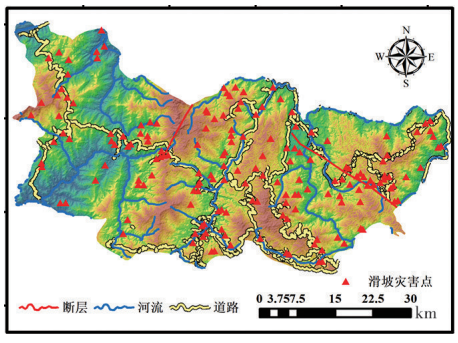


图1　研究区概况图

审图号：GS(2024)0650

2　易发性评价模型构建

2.1　机器学习算法

2.1.1　LightGBM算法

在滑坡地质灾害易发性评价的过程中，构建合适的预测模型是至关重要的。本研究选择了LightGBM[10]（light gradient boosting machine）算法模型作为核心工具之一。LightGBM是一种基于梯度提升框架的机器学习算法，它结合了GOSS（gradient-based one-side sampling）和EFB（exclusive feature bundling）技术，能够在处理大规模数据和具有高维特征的数据集时展现出出色的性能和效率。

LightGBM通过迭代地训练多个弱学习器（通常是决策树），并将它们的结果组合起来以形成一个强学习器，从而实现对目标变量的准确预测。在滑坡易发性评价中，LightGBM能够利用滑坡灾害的历史数据，学习滑坡发生与地质、地形、气候等影响因素之间的复杂关系，进而对新的区域进行滑坡易发性评估。

2.1.2　SVM算法

支持向量机（support vector machine，SVM）[11]作为一种强大的监督学习算法，在滑坡地质灾害易发性评价中也展现出了其独特的优势。SVM通过在高维空间中寻找一个最优超平面，以最大化两类样本之间的间隔，从而实现对目标变量的分类或回归预测。

在滑坡地质灾害易发性评价中，SVM能够充分利用滑坡灾害的历史数据，学习滑坡发生与否与各种地质、地形、气候等影响因素之间的非线性关系。通过核函数的引入，SVM能够处理复杂的非线性问题，使得模型在滑坡易发性评价中具有较高的准确性和泛化能力。此外，SVM在处理高维数据和小样本数据时表现出色，这对于滑坡地质灾害研究来说尤为重要。

2.1.3　RF算法

随机森林（random forest，RF）算法[12]作为一种集成学习方法，在滑坡地质灾害易发性评价中展现出了独特的优势。RF算法通过构建多个决策树并综合其预测结果来提高模型的准确性和稳定性。在滑坡地质灾害研究中，RF算法能够充分利用滑坡灾害的历史数据，学习滑坡发生与否与地质、地形、气候及人类活动等多种影响因素之间的复杂关系。RF算法在处理高维数据、非线性关系和避免过拟合方面表现出色。它能够在不牺牲太多解释性的前提下，通过集成多个弱学习器来构建一个强学习器，从而提高预测的准确性和泛化能力。在滑坡地质灾害易发性评价中，RF算法能够自动处理特征选择问题，即自动选择对滑坡发生有重要影响的因子，而无需进行繁琐的人工特征选择。此外，RF算法对数据的分布要求较低，能够处理不平衡数据，这对于滑坡地质灾害研究来说尤为重要。

2.2　滑坡敏感性因子选取

本研究基于地质、地形和人类活动等[13-16]多方面因素的综合考虑，选取了高程、坡度、坡向、起伏度、曲率、断层距、岩性、NDVI、河流距、TWI、道路距、土体利用类型这12个敏感性因子作为评价的基础。为了获取研究区的敏感性因子数据，采用了多种高分辨率数据源。首先，从Google卫星地图上下载了研究区的卫星影像图，并通过矢量化处理得到了道路和河流的矢量数据。其次，在地理空间信息云（http://www.gscloud.cn/home）上下载了ASTER GDEM 30 m

数字高程模型（digital elevation model，DEM），分辨率为30 m，用于提取高程、坡度、坡向及计算地形湿润指数（TWI）等因子。土地利用类型数据来源于GlobeLand30 2010版，分辨率为30 m。此外，断层及岩性数据则来自于中国国家地质图数据网（http://www.ngac.org.cn/DataSpecial/geomap.html）上下载的1∶500 000地质图，并通过矢量化处理得到。具体敏感性因子分布情况见图2。

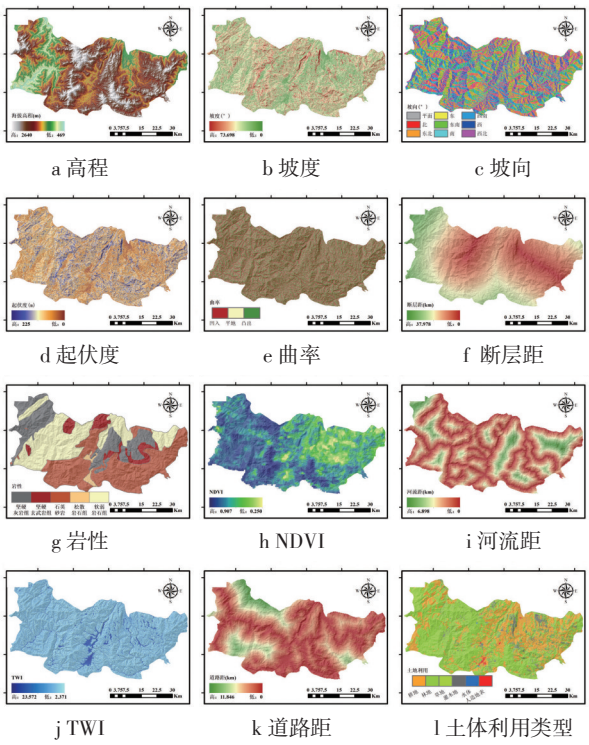


图2 研究区滑坡敏感性因子

审图号：GS(2024)0650

为了构建准确的滑坡地质灾害易发性评价模型，需要对选取的敏感性因子进行相关性分析。相关性分析可以揭示各因子之间的关联程度，从而避免在模型中引入冗余信息[17]。本研究采用了皮尔森相关性分析方法来评估各因子之间的相关性（见图3）。皮尔森相关性系数取值范围在-1~1之间，绝对值越大表示相关性越强。在相关性分析过程中，发现坡度和起伏度之间的相关性值高达0.98，属于强相关关系[18]。若模型中同时引入这两个因子会导致信息冗余，增加模型的计算复杂度，并影响模型的准确性和稳定性[19]。因此，根据相关性分析的结果，需要剔除起伏度这一因子。其余11个因子之间无明显相关性，因此被保留作为模型输入的特征变量。这些因子涵盖了地质、地形和人类活动等多个方面，能够全面反映滑坡地质灾害的易发性特征[20]。

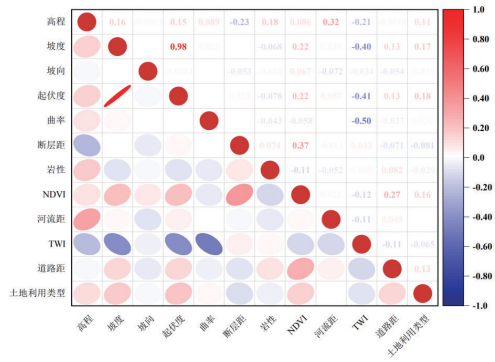


图3 敏感性因子相关性热力图

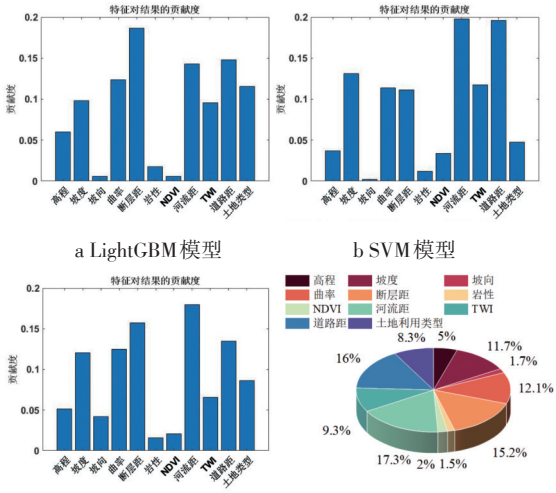
3 滑坡地质灾害易发性评价及区划

针对沧源县展开滑坡地质灾害易发性评价研究，从滑坡样本中随机选择70%样本作为训练样本，剩下30%作为检验集。采用LightGBM、RF和SVM三种机器学习模型进行预测，并对模型结果进行深入的分析。整体而言，LightGBM模型展现出最优的性能，其次是RF模型，SVM模型的表现相对较弱。

3.1 定性分析

从定性分析的角度来看，3个模型在预测研究区东部的滑坡灾害方面显示出良好的一致性。特别是东部地区，作为滑坡灾害的高发区，同样被3个模型预测为高风险区域。其中，LightGBM模型预测的滑坡危险性区域与实际滑坡发生点的一致性极高，从而验证了其在滑坡地质灾害易发性评估中的准确性和可靠性。进一步讨论各敏感性因子的重要性，图4中采用SHAP（shapley additive exPlanations）算法分析的各敏感性因子对模型预测高易发性结果的贡献度，揭示了不同地理和环境因素对边坡失稳的影响程度。图4d展示了各因子的平均贡献度对比，其中河流距离（17.336%）和断层距离（15.181%）主要控制着边坡的稳定性，其次是道路距离（15.958%）、曲率（12.083%）和坡度（11.660%）。这些因素可能通过影响水流、地质稳定性、交通状况和地形特征等方面，同样对滑坡灾害产生显著影响。相比之下，高程（4.956%）、NDVI（2.027%）、岩性（1.527%）和坡向（1.666%）的贡献度相对较小，但在局部地区仍可能具有重要作用。通过TWI（9.283%）和土地利用类型（8.323%）的分析结果进一步揭示，高地形湿度和低植被覆盖区域更易发生滑坡灾害。参照图4，滑坡的分布与道路沿线、断层活动区域及河流两岸高度一致，这与极高危险区域的分布趋势相吻合。综合以上结果，表明道路距离、断层距离和河流距离是影响滑

坡灾害易发性的关键因素，这些因素的分布趋势与极高危险区域的分布高度一致，从而证实了河流侵蚀、断层破裂带和人类活动对滑坡灾害易发性具有显著影响。



a LightGBM模型　　b SVM模型

c RF模型　　d 因子贡献度均值饼状图

图4　滑坡易发性敏感性因子贡献度对比

3.2　定量分析

为了更客观地评价模型的性能，进一步采用ROC曲线及Precision（精确率）、Recall（召回率）、F1_score（*F*1分数）、Accuracy（准确率）和Kappa系数等作为定量评价指标，得到结果见表1。LightGBM模型的AUC值高达0.836，表明其在滑坡地质灾害易发性评价中具有很高的分类准确性。相比之下，RF模型的AUC值为0.773，虽然略低于LightGBM模型，但仍然表现出较好的分类性能。而SVM模型的AUC值仅为0.671，出现假阳性和假阴性较多，有着较多的“误判”，反映了其在处理复杂地质环境、稀疏数据和多因子影响时的局限性。从Precision、Recall、F1_score等指标来看，LightGBM模型在精确率和召回率上均表现最佳，*F*1分数也相对较高，这表明LightGBM模型在预测滑坡地质灾害时，能够较好地平衡假阳性和假阴性的比例，从而提高预测的准确性。*RF*模型在Recall上表现最佳，说明其在识别滑坡地质灾害方面具有较强的敏感性，但在Precision上略低于LightGBM，意味着其预测结果中假阳性的比例稍高。SVM模型在这些指标上的表现均不如前两者，特别是在Accuracy和Kappa系数上，显示了其在分类准确性和一致性上的不足。

综上所述，LightGBM模型在滑坡地质灾害易发性评估中表现突出，其预测的滑坡高风险区域与实际发生的滑坡点紧密对应，同时精确识别了影响滑坡发生的关键因素。即使在数据相对稀缺的研究区域，LightGBM模型仍能发挥出色的预测能力和高准确性，可成为滑坡地质灾害易发性评估的首选方案。

表1　模型的性能比较

模型	LightGBM	SVM	*RF*
Precision	0.813	0.644	0.696
Recall	0.722	0.604	0.765
F1_score	0.765	0.624	0.729
Accuracy	0.765	0.660	0.718
Kappa	0.531	0.315	0.437
AUC	0.836	0.671	0.773

4　结　语

本研究旨在解决滑坡样本稀缺区域的易发性评价难题，选用LightGBM、RF和SVM三种机器学习算法，结合12个影响因子构建评价模型。研究结果显示，LightGBM模型预测性能最优，其滑坡危险区划定结果与实际灾害点高度吻合，在研究区东部高易发区表现尤为突出；该模型的AUC值达0.836，显著高于RF和SVM。影响因子重要性分析表明，河流距（17.336%）和断层距（15.181%）贡献最大，道路距、曲率与坡度次之，揭示出河流侵蚀、地质构造和人类工程活动是影响滑坡发生的关键要素。该研究为小样本条件下的滑坡风险识别提供了有效方法与科学依据，具备推广价值。

参考文献

[1] 向喜琼. 区域滑坡地质灾害危险性评价与风险管理[D]. 成都:成都理工大学,2005

[2] 黄润秋. 20世纪以来中国的大型滑坡及其发生机制[J]. 岩石力学与工程学报,2007,25(3):433-454

[3] 谭彬,刘国栋,余安亮,等. 基于AHP-INF耦合模型的察雅县地质灾害易发性评价[J]. 地理空间信息,2023,21(11):15-19

[4] 罗相涛. 九寨沟漳扎地震区滑坡易发性对比分析[J]. 地理空间信息,2023,21(6):103-107

[5] Tien B D, Tuan T A, Klempe H, et al. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree[J]. Landslides,2016,13(2):361-378

[6] 林琴,郭永刚,吴升杰,等. 基于梯度提升的优化集成机器学习算法对滑坡易发性评价:以雅鲁藏布江与尼洋河两岸为例[J]. 西北地质,2024,57(1): 12-22

[7] 黄发明,胡松雁,闫学涯,等. 基于机器学习的滑坡易发性预测建模及其主控因子识别[J]. 地质科技通报,2022,41(2):79-90

（下转第73页）