



DOI:10.12404/j.issn.1671-1815.2403254

引用格式:袁旭山,刘京会,黄龙生,等.基于机器学习模型的新源县滑坡易发性评估[J].科学技术与工程,2025,25(5):1815-1826.

Yuan Xushan, Liu Jinghui, Huang Longsheng, et al. Assessment of landslide susceptibility in xinyuan county based on machine learning models[J]. Science Technology and Engineering, 2025, 25(5): 1815-1826.

基于机器学习模型的新源县滑坡易发性评估

袁旭山,刘京会*,黄龙生,李鑫旭

(防灾科技学院应急管理学院,三河 065201)

摘要 滑坡灾害对居民生活和社会经济构成严重威胁。以新源县为研究区,选取17个滑坡影响因子作为初始因子集,通过多重共线性分析,筛选10个滑坡因子并构建研究区滑坡易发性评估指标体系。分别基于逻辑回归(logistic regression, LR)、支持向量机(support vector machine, SVM)和随机森林(random forest, RF)3种典型模型进行滑坡易发性评估。利用受试者操作特征(receiver operating characteristics, ROC)曲线下的面积(area under curve, AUC)、滑坡比及野外实地考察对模型评估结果进行对比验证分析。结果显示:低易发区主要集中在巩乃斯河谷平原地区,这些区域地势平坦,滑坡易发性相对较低。高易发区主要集中于巩乃斯河谷北部、阿吾拉勒山丘陵区,以及南部伊什基里克山和那拉提山分水岭两侧、恰甫河以南区域,地形复杂多变,滑坡易发性较高。在3种评估模型中,SVM模型表现最佳,其AUC高达0.985,说明该模型在滑坡易发性评估中的准确性较高。此外,SVM模型评估得到的高易发区滑坡点密度大,占比达到86%,进一步验证了其在滑坡易发性评估中的有效性。综合以上结果,SVM模型在新源县滑坡易发性评估中的合理性优于其他两种算法,能够为该地区的滑坡防治提供科学的理论依据和参考。

关键词 逻辑回归;支持向量机;随机森林;滑坡易发性;新源县

中图分类号 P642; **文献标志码** A

Assessment of Landslide Susceptibility in Xinyuan County Based on Machine Learning Models

YUAN Xu-shan, LIU Jing-hui*, HUANG Long-sheng, LI Xin-xu

(School of Emergency Management, Institute of Disaster Prevention, Sanhe 065201, China)

[Abstract] Landslide disasters pose a serious threat to residents' lives and socio-economic development. Taking Xinyuan County as the study area, 17 landslide influencing factors were selected as the initial factor set. Through multiple collinearity analysis, 10 landslide factors were screened and an evaluation index system for landslide susceptibility in the study area was constructed. The landslide susceptibility was then evaluated based on three typical models: logistic regression (LR), support vector machine (SVM), and random forest (RF). The evaluation results of the model were compared and validated using the area under curve (AUC), landslide ratio, and field investigation under the receiver operating characteristics (ROC) curve. The results show that low-susceptibility areas are mainly concentrated in the valley plain of the Kongnais River, where the terrain is flat and the landslide susceptibility is relatively low. High-susceptibility areas are mainly located in the northern part of the Kongnais River valley, the Awulale hilly area, and the watersheds on both sides of the southern Yishikelike Mountains and Nalati Mountains, as well as the area south of the Qiafu River, where the terrain is complex and varied, leading to a higher susceptibility to landslides. Among the three evaluation models, the SVM model performs the best, with an AUC value of up to 0.985, indicating its high accuracy in landslide susceptibility assessment. Furthermore, the high-susceptibility areas identified by the SVM model have a high density of landslide points, accounting for 86% of the total, further validating its effectiveness in landslide susceptibility assessment. Based on the above results, the SVM model is more reasonable than the other two algorithms in assessing landslide susceptibility in Xinyuan County, providing a scientific theoretical basis and reference for landslide prevention and control in the region.

[Keywords] logistic regression; support vector machine; random forest; landslide susceptibility; Xinyuan County

收稿日期:2024-05-03 修订日期:2024-12-23

基金项目:第三次新疆综合科学考察项目(2022xjkk0600)

第一作者:袁旭山(2000—),男,汉族,山东寿光人,硕士研究生。研究方向:灾害监测与评估。E-mail:2505475983@qq.com。

*通信作者:刘京会(1975—),女,汉族,河北保定人,博士,副教授。研究方向:灾害监测与评估、灾害遥感与GIS应用。E-mail:liujh@cidp.edu.cn。

滑坡,作为一种频发且极具破坏性的地质灾害,其对人类社会的影响不容忽视^[1]。中国地域辽阔,多山地貌显著,这导致许多地区频繁受到滑坡灾害的严重影响^[2]。近年来,滑坡对环境、居民建筑和工业设施的威胁日益加重^[3],对居民的生命和财产安全构成了严重威胁,给国家和社会带来了巨大损失^[4]。迫切需要一种有效的方法来减少滑坡带来的损害。因此,开展科学、准确的滑坡危险性评估对于制定防灾减灾措施具有重要的指导意义。

滑坡易发性评估旨在对滑坡灾害发生的可能性进行科学的评估,并提供有关滑坡的时空分布和发生概率的信息。其目的是回答哪些地区最容危险生滑坡这样一个空间概率问题^[5],通过对滑坡危险性的评估,可以了解不同地区发生滑坡灾害的潜在风险程度。评估结果可以帮助决策者和相关部门制定有效的滑坡灾害风险管理措施,包括预防、减灾和应急管理方面的措施^[6]。此外,滑坡危险性评估结果还可以为国土空间规划布局提供参考。了解不同地区的滑坡危险性可以帮助规划者避免在高风险区域进行敏感的土地利用规划,从而减少滑坡灾害对人类活动和财产造成的风险和损失^[7]。滑坡灾害易发性评估是通过选择适当的指标体系、结合历史灾害数据,并采用特定的评估模型进行计算,最终得到评估区域划分图的过程。

目前广泛使用的区域滑坡灾害评估模型包括:分类识别模型,如卷积神经网络^[8]、BP神经网络^[9];统计预测模型,如证据权^[10]、确定性系数^[11]、信息量^[12]。这些模型各自具有一些优点和缺点。分类识别模型,通过深度训练已知样本,展现了对未知样本进行精准分类和预测的能力。这些模型显著的特点在于其高度的并行处理能力和对非线性全局效应的有效应对。即使在面对复杂的参数关系,尤其是在未知变量与自变量间没有明确函数关系的情况下,它们依然能够出色地处理这些非线性映射关系,确保分类或预测的准确性。然而,在处理滑坡易发性评估过程中的存在的不确定性问题时,当前的分类识别模型确实存在一定的局限性,其表现并不理想。这类模型通过深入系统地分析历史地质灾害数据,成功识别出导致过去不稳定地质过程的关键综合因素参数。随后,这些参数被巧妙地应用于那些具有相似地质条件但尚未发生滑坡的地区,以实现定量或半定量的风险评估。然而,需要指出的是,滑坡易发性评估模型的精确性在很大程度上依赖于所使用的数据的完整性和详细程度。因此,确保数据的质量是这类模型能否发挥其应有功能的关键。

近年来,机器学习模型在滑坡灾害易发性评估中得到广泛的应用^[13]。提高了评估的准确性和效率,使得评估结果更为可靠,而且能够更好地处理不确定性问题。通过结合多源数据和综合模型的结果,可以提供更全面、客观的滑坡易发性评估,为滑坡灾害的防治和管理提供了重要的科学依据^[13]。因此,现选择新源县作为研究区域,分别基于逻辑回归(logistic regression, LR)支持向量机(support vector machines, SVM)和随机森林(random forest, RF)3种典型机器学习算法获取该地区的滑坡易发性评估结果,并使用不同精度评估指标对结果进行比较分析,以确定最佳评估算法和结果。所得到的最优评估结果将为当地政府部门提供更准确、可靠的防灾减灾参考依据。

1 研究区概况与数据来源

1.1 研究区概况

如图1所示,新源县地处新疆维吾尔自治区西部,坐落于天山深处的伊犁河谷东部,地理位置极为关键,具体坐标横跨东经 $82^{\circ}28' \sim 84^{\circ}57'$,北纬 $43^{\circ}01' \sim 43^{\circ}40'$,总面积达到 $7\,580.62\text{ km}^2$ 。该县交通网络发达,217、218国道以及316省道在此交织,为连接中亚、北疆及南疆的重要枢纽。新源县的地形特色鲜明,四周三面被群山环绕,而西部则是一片较为宽广的平原,整体呈现出东高西低的自然地形趋势。县域内山峦起伏,山地地形占据总面积的73.2%,为其增添了显著的山地特色。然而,由于独特的气候、地貌、地质条件以及人类活动的影响,新源县成为了新疆地区地质灾害最为严重的县市之一。该县地质灾害类型多样,频发且影响深远,受灾范围广泛,危害日益加剧。在全球气候变暖及人类工程活动不断扩大的背景下,新源县的地质灾害呈现出更为集中、频发的态势。新源县地质灾害的范围进一步扩大,这不仅加剧了灾害的严重程度,也导致了经济损失的不断攀升。这种局面给新源县各族人民的生命安全和财产安全带来了日益严峻的挑战^[14]。

1.2 数据来源

为了评估新源县的滑坡灾害风险情况,进行了广泛的滑坡灾害资料搜集工作,并准备了各种影响因子数据,获取了445个滑坡点的地理位置信息,涵盖了整个滑坡影响区内所有可调查到的滑坡灾害致灾因子。建立了研究去滑坡空间数据库,并对新源县滑坡灾害的孕灾环境进行了深入分析。

滑坡的发育受到内在和外在因素的共同影响。内在因素包括岩土体的结构和性质等。外在因素包

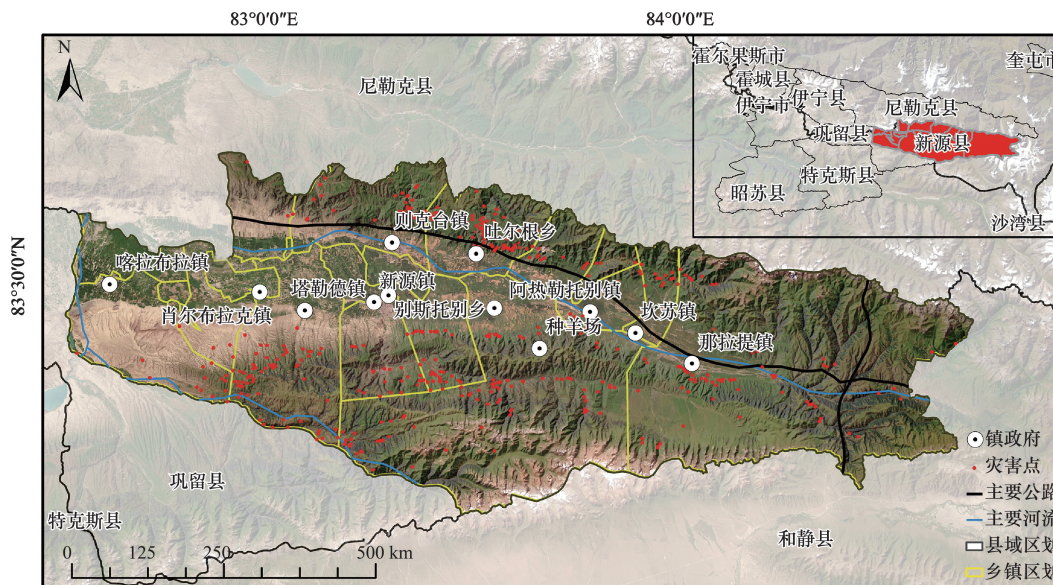


图1 研究区分布图

Fig. 1 Distribution map of the research area

括降雨、人类活动等,而地形特征被认为是滑坡形成的主要影响因素之一^[15]。李萍等^[16]的研究在评估永德县地质灾害危险性时,综合考量了地理环境、地质环境以及人类工程活动3个关键维度,并精心选取了14项指标因子作为评估依据。这些全面的指标选择为永德县地质灾害易发性的准确评估提供了坚实的基础。李利峰等^[17]利用多种机器学习模型对原始和改进影响因子进行比较,最终确定水系密度和公路密度相较于距水系距离和公路距离在滑坡危险性评估中表现更优。刘帅等^[18]构建

了一个包含9个评估指标的体系,用于对不同降雨工况条件下的崩滑地质灾害进行易发性评估。采用了一系列综合性的指标,包括工程地质岩组的特性、地貌类型的分类、地面的高程变化、地形的坡度、坡向以及地表的曲率等自然地理因素。同时,还考虑了人类生成社会活动对地质条件的影响,如土地利用的类型、与河流的距离以及与潜在断层区域的距离等。基于新源县的地质环境特征和历史滑坡数据,以及对滑坡发生机理的理解,选择了17个滑坡影响因子作为初始因子集,如表1所示。

表1 滑坡危险性评估数据及来源

Table 1 Landslide hazard assessment data and sources

数据获取	数据来源	滑坡影响因子
地质灾害点分布数据	地理遥感生态网 http://www.gisrs.cn	滑坡点
		高程(elevation)
		坡度(slope)
		坡向(slope orientation)
30 m 精度 DEM 数字高程数据	地理空间数据云 http://www.gscloud.cn	平面曲率(plan curvature)
		剖面曲率(profile curvature)
		地形起伏度(relief)
		地形湿度指数(terrain humidity index)
		水流强度指数(stream power index)
30 m 土地覆盖数据	全球土壤覆盖数据库 http://www.globallandcover.com	地貌类型(landform type)
		土地覆盖类型(land cover type)
岩性数据	中国科学院资源环境科学数据中心 http://www.resdc.cn	归一化植被指数(normalized vegetation index)
30年降雨数据	国家青藏高原科学数据中心 https://data.tpdc.ac.cn	岩性(lithology)
中国地震动峰值加速度区划图	国家地震科学数据中心	年降雨量(annual rainfall)
1:400万活动断层数据	https://data.earthquake.cn/index.html	地震动峰值加速度(earthquake peak acceleration)
1:25万水系(线)数据	全国地理信息资源目录服务系统	距断层距离(distance from fault)
矿场分布数据	https://www.webmap.cn	距河流距离(distance from river)
		矿场密度(mine density)

2 研究方法

2.1 技术路线

通过相关性分析和重要度分析构建了滑坡易发性评估指标体系,并使用逻辑回归、支持向量机和随机森林等机器学习模型进行评估。通过受试者操作特征(receiver operating characteristics, ROC)曲线验证、滑坡比和实地考察验证,对模型结果进行了精确性和可靠性的分析和比较。

2.2 评估模型及验证

2.2.1 逻辑回归模型(LR)

逻辑回归特别适用于二分类问题,但也可扩展到多分类场景。通过描述和验证滑坡或非滑坡和滑坡影响因子之间的关系进行滑坡易发性评估^[19]。在滑坡易发性评估中,为构建有效的模型,首先使用 ArcGIS 软件的随机创建点工具,在滑坡点周边 1 km 范围内随机生成等量的非滑坡点。接着,将滑坡点和非滑坡点组合成一个总样本集。为了训练滑坡

危险性评估模型,首先从总样本集中选取了 70% 的滑坡点及其相应数量的非滑坡点作为训练样本。剩余的 30% 滑坡点和相应非滑坡点则用作测试样本,以验证模型的准确程度。在建模过程中,将其作为关键变量并作为模型的输入自变量,以提高预测精度。同时,将滑坡的发生情况设为分类因变量,其中 1 代表滑坡发生,0 代表滑坡未发生。滑坡发生概率被设为 $P, P \in [0, 1]$ 。 P 的表达式为

$$P = P(Y|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}} \quad (1)$$

式(1)中: Y 为结果变量(滑坡或非滑坡); X_1, X_2, \dots, X_n 为滑坡影响因子的预测变量, n 为滑坡影响因子的个数; e 为自然对数的底数; β_0 为截距条件; $\beta_1, \beta_2, \dots, \beta_n$ 为回归系数^[20]。

2.2.2 支持向量机模型(SVM)

Vapnik^[21] 提出支持向量机算法,这是一种机器学习算法,其核心建立在统计学习理论上,通过对数

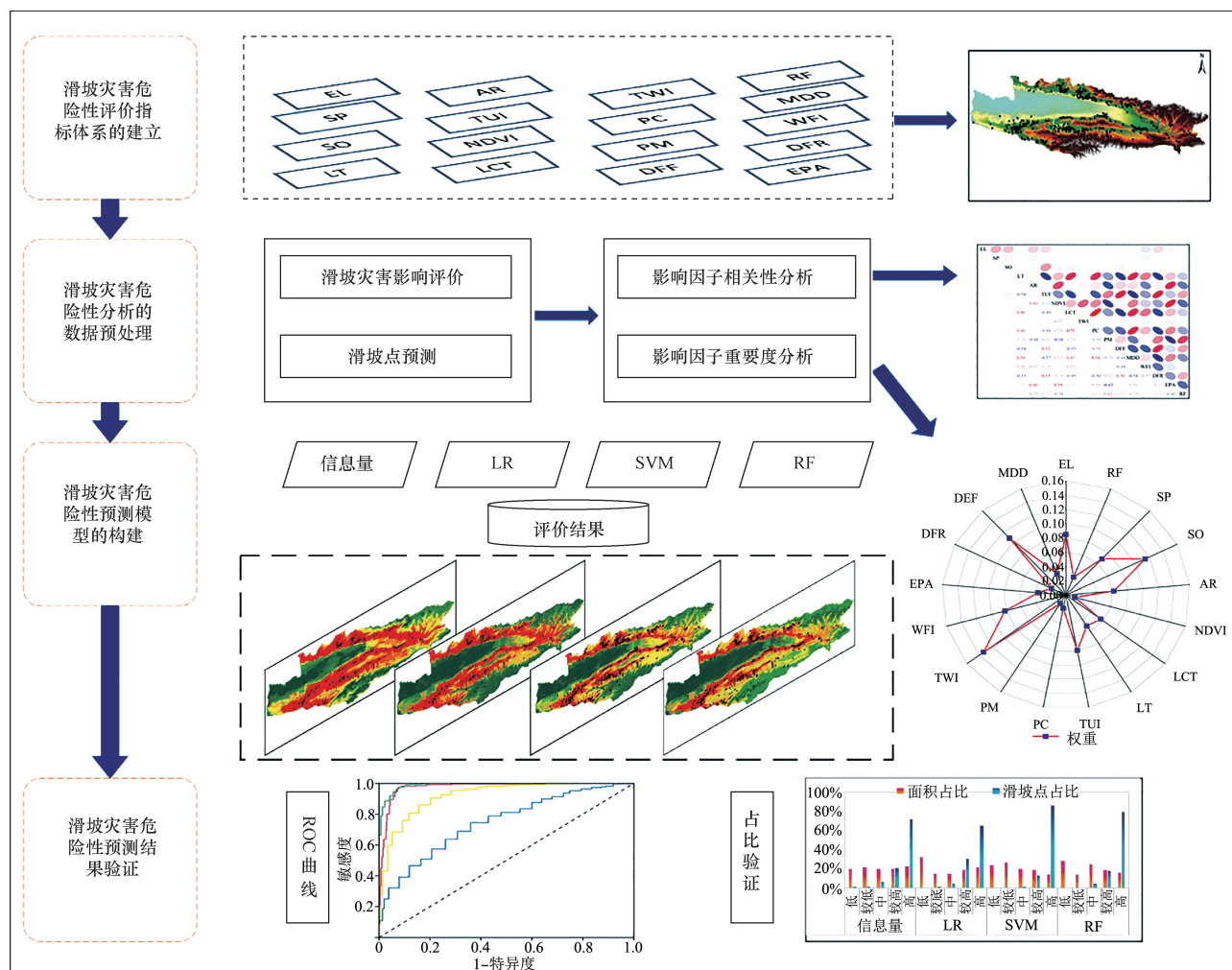


图2 技术路线图

Fig. 2 Technical Roadmap

据的统计分析来学习和改进模型性能。令 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ 为 n 个滑坡影响因子的向量, $\mathbf{Y}_i = (Y_1, Y_2)$ 为滑坡点与非滑坡点的向量。SVM 通过寻找最优超平面将两类数据正确分开,并保持其间隔达到最大值,其最优分离超平面可以通过求解分类函数得到形式^[22]如下。

$$f(\mathbf{X}) = \text{sign} \left[\sum_{i=1}^n a_i Y_i k(\mathbf{X}, \mathbf{X}_i) + c \right] \quad (2)$$

式(2)中: c 为距超平面原点的偏移量; n 为滑坡影响因子个数; a_i 为正实常数; $k(\mathbf{X}, \mathbf{X}_i)$ 为核函数,可以是线性、多项式、径向基函数,也可以是 Sigmoid^[23]。从预先建立的总样本集中随机选取了70%的滑坡点和相同数量的非滑坡点作为训练样本,剩余的30%滑坡点和非滑坡点则作为测试样本。随后,使用 MATLAB 软件,基于支持向量机(SVM)模型对这些样本进行训练,以获取滑坡易发性的评估结果。

2.2.3 随机森林模型(RF)

随机森林是一种集成学习模型,它结合了多个决策树分类器的预测结果,通过整合这些基本单元(决策树)的输出来提升整体的分类性能。这种机器学习算法的工作原理是构建多个决策树,并在每个树的构建过程中随机选择特征和样本子集进行训练。这种方法旨在减少过拟合现象,并提升模型的泛化能力。该模型的步骤如下:先通过无放回的从原始训练样本中随机抽取 n 个训练样本构建决策树模型,然后在随机抽取 m 个特征子集,并筛选出最优特征子集对其进行生长,最后经过决策树投票获得最终结果^[24]。随机森林模型可用公式表示为

$$Z(x) = \arg \max_m \sum_{i=1}^n I[Y_i(x) = U] \quad (3)$$

式(3)中: $Z(x)$ 为随机森林模型; $Y_i(x)$ 为单个决策树模型; U 为输出变量; $I(\cdot)$ 为显函数; m 为特征子集个数。通过 MATLAB 软件,调用随机森林模型,用于预测滑坡发生的可能性,从而进行滑坡易发性评估。在设置参数时,将随机森林树的深度设置为100,从总样本集中抽取了70%的滑坡点和与之等量的非滑坡点作为训练数据,而剩余的30%则作为独立的测试样本。通过这种方式,确保模型能够基于这些训练样本进行学习,并在后续使用测试样本评估其性能,再将学习后的模型进行测试样本。

2.2.4 模型精度检验

ROC 曲线是评估预测模型性能的重要工具。它基于混淆矩阵的原理,通过设定不同的二分类阈值,以灵敏性为纵坐标,1 - 特异性为横坐标绘制而

成。ROC 曲线直观展示了模型在不同阈值下对正例和负例的区分能力。当 ROC 曲线越接近左上角时,意味着模型在保持低假正率的同时,具有较高的真正率,这反映了模型的高准确性。通过观察 ROC 曲线的形状和位置,可以直接评估学习器的性能,判断其分类效果的优劣。因此,ROC 曲线不仅是一个简单直观的评估工具,也是分析模型性能的重要参考依据^[25]。

$$\text{真阳性率:TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

式(4)中:TP为被正确地判定为正例的正例样本数量;FN为被错误地判定为负例的正例样本数量。

$$\text{假阳性率:FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (5)$$

式(5)中:FP为被错误地判定为正例的负例样本数量;TN为被正确地判定为负例的负例样本数量。

3 结果与分析

3.1 滑坡影响因子的筛选

共选取了17个影响因子(包括坡度、坡向、高程等)作为初始滑坡影响因子。为了评估这些因子的相关性和重要性,采用皮尔逊相关性系数进行分析,当相关系数小于0.5时,说明影响因子之间的相关性较小^[26]。通过反复训练模型,并结合相关性图(图3)和重要度分布图(图4),对这17个初始影响因子进行筛选,识别对模型性能具有关键作用的重要因子。在筛选过程中,剔除了相关性较高但重要程度较低的因素,最终选取了10个滑坡影响因子,包括高程、坡度、坡向、地貌类型、土地覆盖类型、岩性、归一化植被指数(NDVI)、年均降水量、水流强度指数以及地形湿度指数(其中水流强度指数 $\text{WFI} = A_s \tan \beta$;地形湿度指数 $\text{TWI} = \ln \frac{A_s}{\tan \beta}$,WFI为水流强度指数,TWI为地形湿度指数, A_s 为单位汇水面积, β 为坡度)。基于筛选后的影响因子分析滑坡点和影响因子之间的关系,由图5(a)~图5(d)可知,滑坡主要集中在高程1000~2500m,坡度 $10^\circ \sim 34^\circ$,随着坡向和NDVI的增加,滑坡数量逐渐增加;由图5(e)~图5(d)可以看出,滑坡主要分布于森林、草原及丘陵、低山的地貌环境中,岩性种类以板岩为主,同时,受降水影响,在150~300mm滑坡数较多;从图5(i)、图5(j)可以看出,对土壤的侵蚀成度和水流强度也会不同程度的影响滑坡的发生。此外,为了保证因子之间具有统一的栅格单元大小,还利用重采样工具将栅格单元的分辨率调整为 $30\text{ m} \times 30\text{ m}$,并根据滑坡灾害点的分布规律对各评估因子进行建模(图6)。

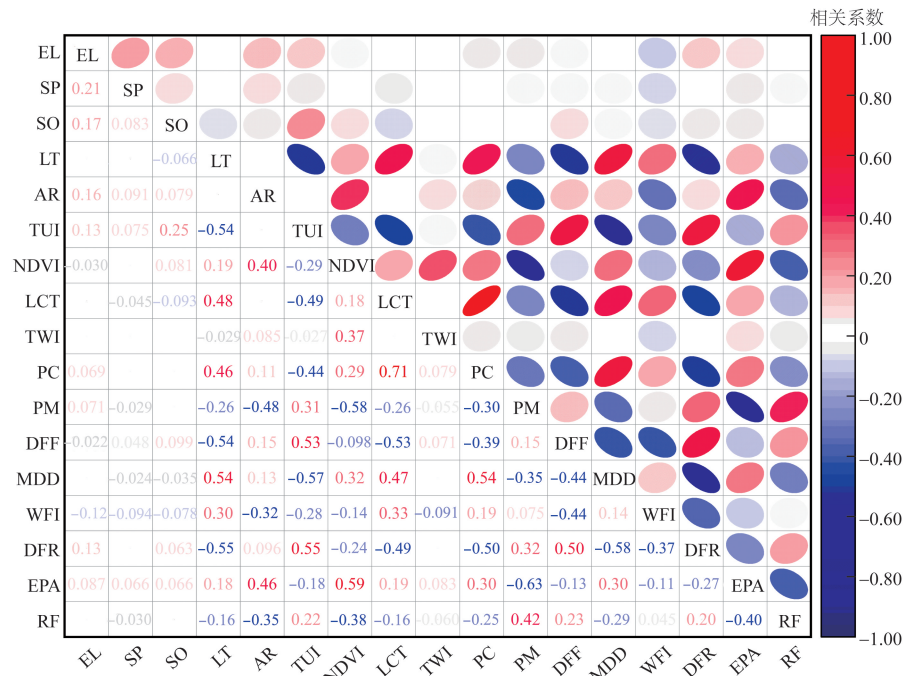


图3 相关性分析结果

Fig. 3 Correlation analysis results

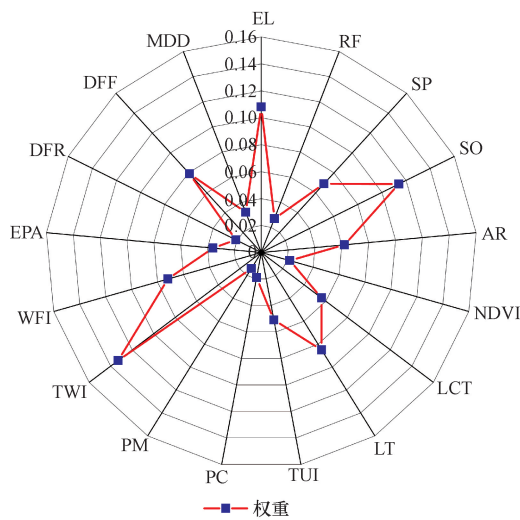


图4 重要性分析结果

Fig. 4 Importance analysis result

3.2 滑坡易发性结果分析

在滑坡分布及其影响因子研究中,选择 30m 分辨率的栅格形式进行统一处理,以确保数据的标准化和一致性。研究区域共覆盖了 8 404 809 个栅格,其中涉及滑坡的栅格数为 445 个。为检验滑坡危险性区划图的精确性,借助 ArcGIS 软件在滑坡点周边 1 km 的区域内随机抽取了 445 个非滑坡点,与既有的滑坡点共同汇编成全面的总样本集,从这一总样本集中挑选出训练样本和验证样本。具体而言,约 70% 的滑坡点和非滑坡点(即 311 个)被用作训练样本,以测试 3 个不同模型的效果;而剩余的 30%

(即 134 个)则作为验证样本,用于评估模型的预测能力。

将 3 个模型得到的滑坡易发性评估图,利用 ArcGIS 的自然断点分级法将其重分类为 5 个不同的区间,确定了 5 个危险等级:低、较低、中、较高和高,并绘制了滑坡易发性结果图(图 7)。与 LR 和 RF 模型相比,SVM 模型在极高和高易发区的估算面积占比更小,这表明 SVM 在识别高易发性地区方面具有更高的精度。这是因为 SVM 能够更好地捕捉数据中的非线性关系,从而提高了对滑坡易发性的预测能力。通过观察 3 种模型绘制的滑坡易发性图,可以观察到滑坡灾害的总体分布趋势如下。

(1) 高易发区和较高易发区主要聚集在几个关键区域,包括巩乃斯河谷的北部、阿吾拉勒山的中低山丘陵地带,以及南部伊什基里克山和那拉提山分水岭的两侧,尤其是恰甫河以南的广泛区域。这些区域地形崎岖、起伏大,地质构造环境复杂,例如南部的伊什基里克山东部和那拉提山北坡具有明显的高程差异性和坡度的多变性,这些特征容易导致土壤失稳,增加了滑坡发生的可能性。其中,库尔德宁山东麓至阔克乔克山北麓滑坡发生的可能性高。巩乃斯河谷的北部、阿吾拉勒山的中低山丘陵地带由于人类工程活动的频繁影响,地质稳定性面临严峻挑战。如在河谷边缘推进农业水利工程、牧业的繁荣发展、公路交通网络的扩张以及矿山资源的开采等人为因素,无疑加剧了这些地区的地质脆

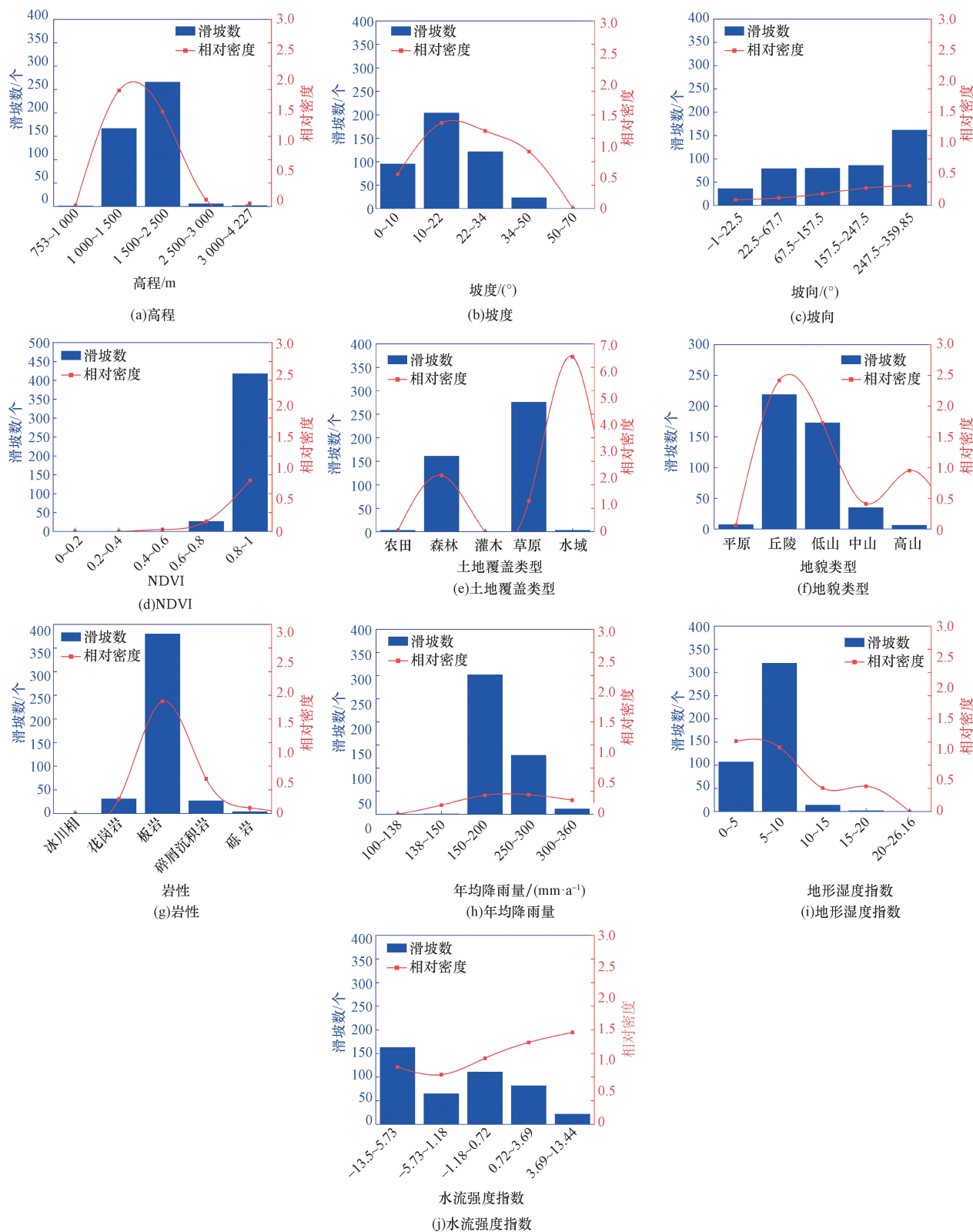


图5 滑坡影响因子分级滑坡数及相对密度统计

Fig.5 Classification of landslide impact factors

弱性。其中,位于阿吾拉勒山南坡的吐尔根乡北部地质复杂,显著特征为断层破碎带广泛分布,同时覆盖有深厚的黄土层,容易诱发滑坡灾害。

(2)中危险区主要环绕新源县北部阿吾拉勒山高危危险区外围,具体涵盖铁木里克勒加克布拉克、吐尔根萨依上游以及坎苏沟上下游地带。在县境南

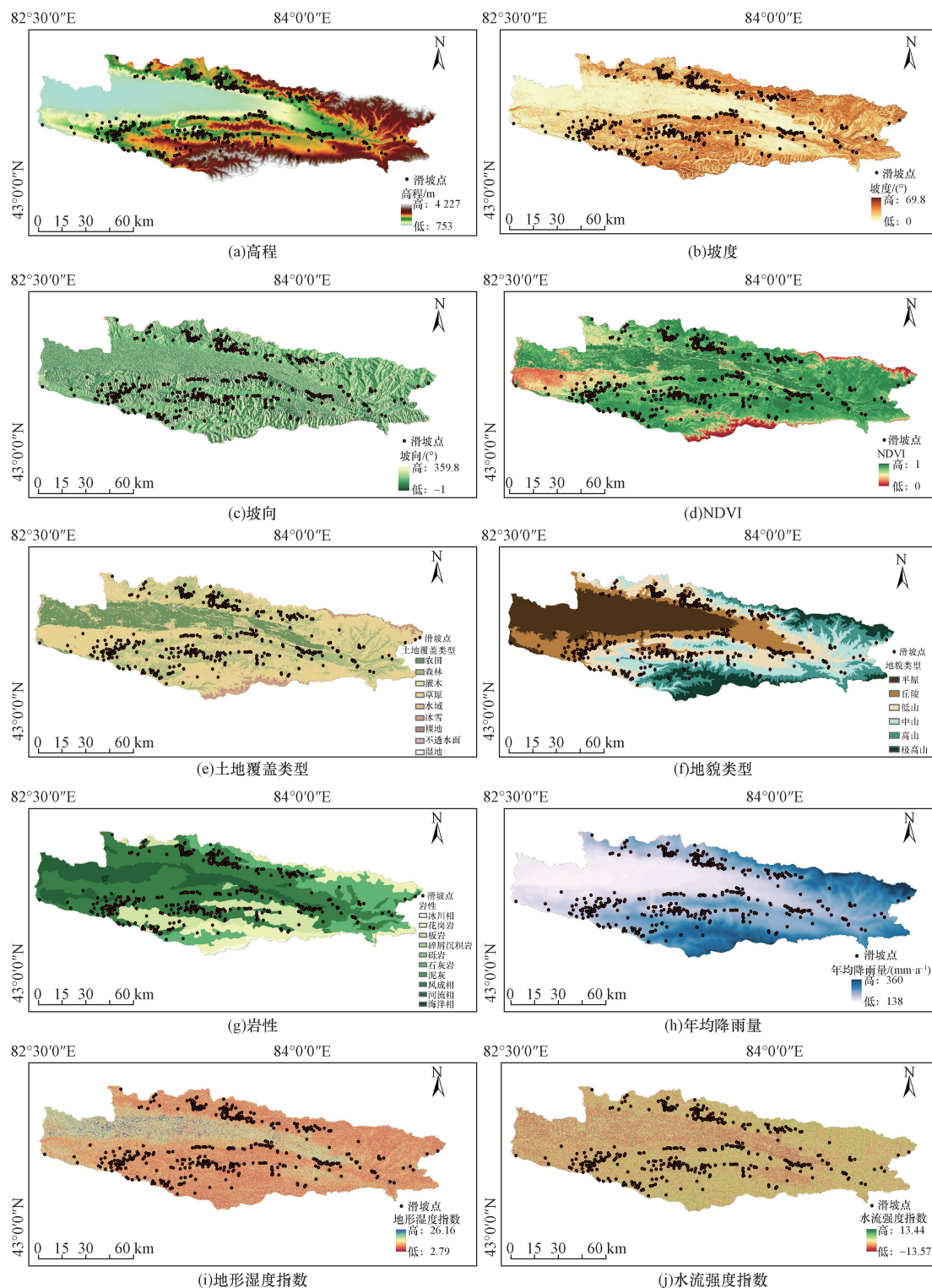


图6 滑坡影响因子

Fig. 6 Landslide impact factors

部,除了伊什基里克山和那拉提山分水岭两侧及恰甫河以南的高易发区外,其外围还分布着一些中易发区,包括塔勒德萨依、阿提别克萨依、吉尔尕朗河上游北岸的那孜巴音塔勒地带,以及拉斯台沟出口处等区域。尽管这些区域的滑坡风险相对较低,

对潜在受灾对象的威胁较小,但仍需保持警惕,因为它们具备地质灾害的潜在诱发条件。这些地带的地质环境条件较为复杂,地质现象相对活跃,但总体上地质灾害发生频率不高。特别是它们的地形特征,陡峭且地势变化显著,成为影响滑坡易发性

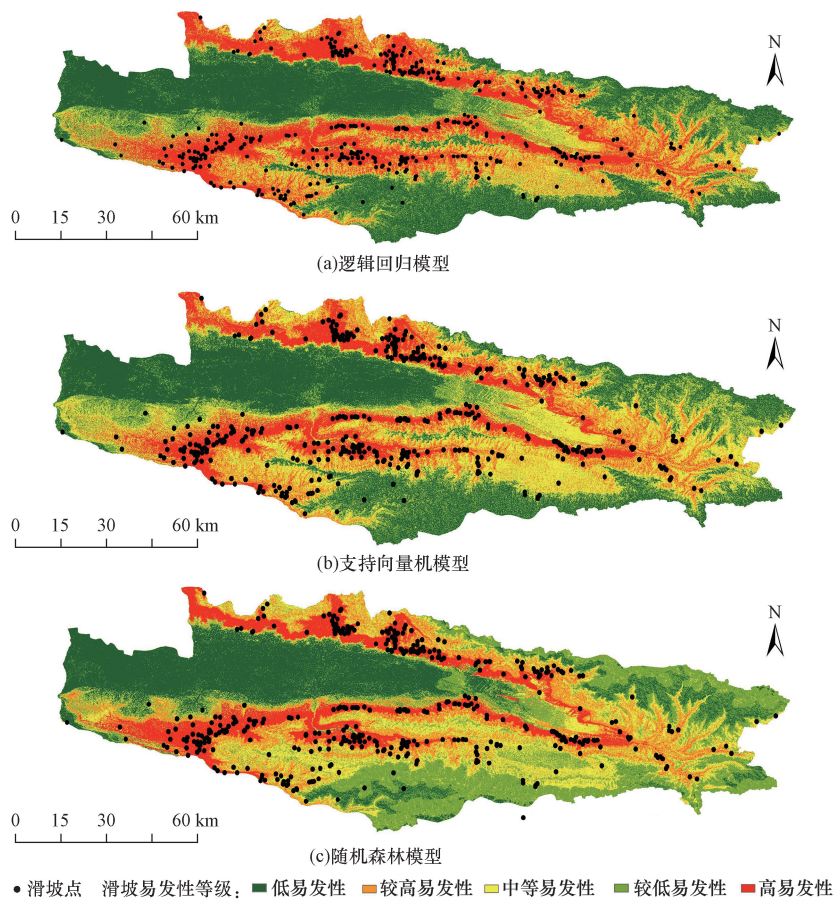


图7 滑坡易发性等级分区

Fig. 7 Landslide hazard level zoning

的重要因素之一。

(3) 低易发区和较低易发区主要分布于县境南部的高中山区和低山丘陵区,尤其是林带以上的区域。这一地区属于巩乃斯河谷平原,地势平坦开阔,缺乏导致滑坡发生和发育的基础环境条件,因此滑坡灾害相对较少。

具体而言,高易发区主要聚集在则克台镇、吐尔根乡、阿热勒托别镇、坎苏镇北部、别斯托别乡、阿勒玛勒镇南部、那拉提镇、新源镇中部、肖尔布拉克镇东南部以及塔勒德镇西南部等地。这些地区滑坡风险显著,与前人研究结果相吻合^[27],是滑坡灾害的重点防范区域。而低易发区则广泛分布在肖尔布拉克乡的中北部、喀拉布拉镇、塔勒德镇北部、新源镇北部、别斯托别乡北部以及阿勒玛勒镇南部等地。研究表明,滑坡的分布与地形坡度、地面高程等地形因素密切相关。坡度是影响滑坡启动的关键因素,过陡或过缓的坡度都可能影响滑坡的发生。此外,降雨量的变化也会对滑坡灾害的形成产生潜在影响。因此,在讨论区域滑坡灾害易发性的影响因素时,需要综合考虑地形因素和气象因素。对于滑坡原因的具体分析,可以结合滑坡影响因子的重要性排序结果进行深入探

讨,以更准确地评估不同区域的滑坡风险,并采取相应的防范措施。

3.3 评估精度对比分析

为了对模型评估结果的准确性进行量化分析,采用统计方法计算了模型测试样本的曲线下的面积(area under curve, AUC),并在 SPSS Pro 中绘制了3种模型的 ROC 曲线(图8)。通过比较图8中的 ROC 曲线表现,

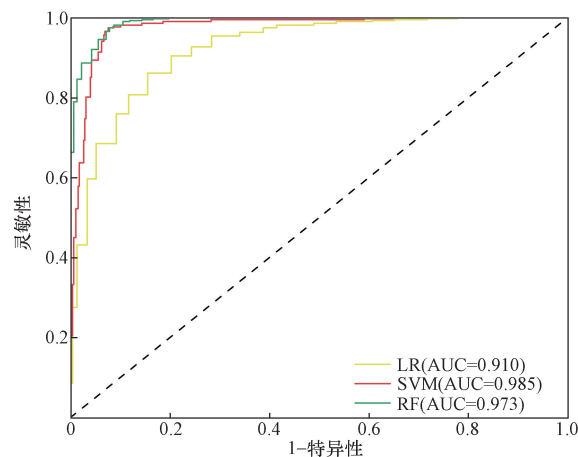


图8 测试样本 ROC 曲线

Fig. 8 ROC curve of test sample

发现在使用相同数据集的条件下,LR、SVM 和 RF 模型的接收者特征曲线下面积分别为 0.910、0.985 和 0.973。总体而言,使用新源县滑坡灾害数据集构建的 3 种机器学习模型都取得了令人满意的效果,其 AUC 均超过 0.9,说明这 3 种机器学习模型都适用于新源县滑坡灾害数据集。在新源县滑坡易发性评估中,SVM 模型在各项性能指标中表现最佳。

在寻求最优滑坡易发性评估模型并确保其合理性具有可比性的过程中,引入了滑坡比这一指标。通过计算易发性等级面积占比与实际滑坡点占比的比值,能够更准确地度量各级易发区内实际滑坡点的分布情况,从而为模型优化提供有力依据。滑坡比定义为实际滑坡点落入各级易发区的面积百分比与各级易发区面积占总面积的百分比之比。根据表 2 的数据,可以得知 LR、SVM、RF 模型在高易发区分级面积占比分别为 21%、14%、16%。检验结果显示,在 LR、SVM、RF 模型的评估结果中,分别有 65%、86%、79% 的滑坡点落入高易发等级区域。因此,根据滑坡比由高到低的顺序,3 种模型的排列顺序为 SVM、RF、LR。这说明 SVM 模型在新源县滑坡易发性评估结果的合理性方面优于其他 3 种算法。

表 2 易发性分区合理性检验结果

Table 2 Results of reasonableness test for hazard zoning

评估模型	易发性等级	分级面积/km ²	滑坡点/个	面积占比/%	滑坡点占比/%
LR	低	238 783	0	32	0
	较低	113 281	4	15	1
	中	110 499	17	15	4
	较高	134 198	133	18	30
	高	159 576	291	21	65
SVM	低	175 326	0	23	0
	较低	195 903	2	26	0
	中	146 783	4	19	1
	较高	132 753	57	18	13
	高	107 376	382	14	86
RF	低	213 721	0	28	0
	较低	106 360	2	14	0
	中	181 433	16	24	4
	较高	136 549	77	18	17
	高	118 335	350	16	79

3.4 野外考察验证分析

如图 9 所示,根据野外实地考察,发现新源县那拉提镇东部那拉提山存在严重的滑坡隐患问题。图 9(d)为那拉提山坡的一个滑坡隐患点,可能会对当地游牧居民生命安全造成一定威胁,为了提醒当

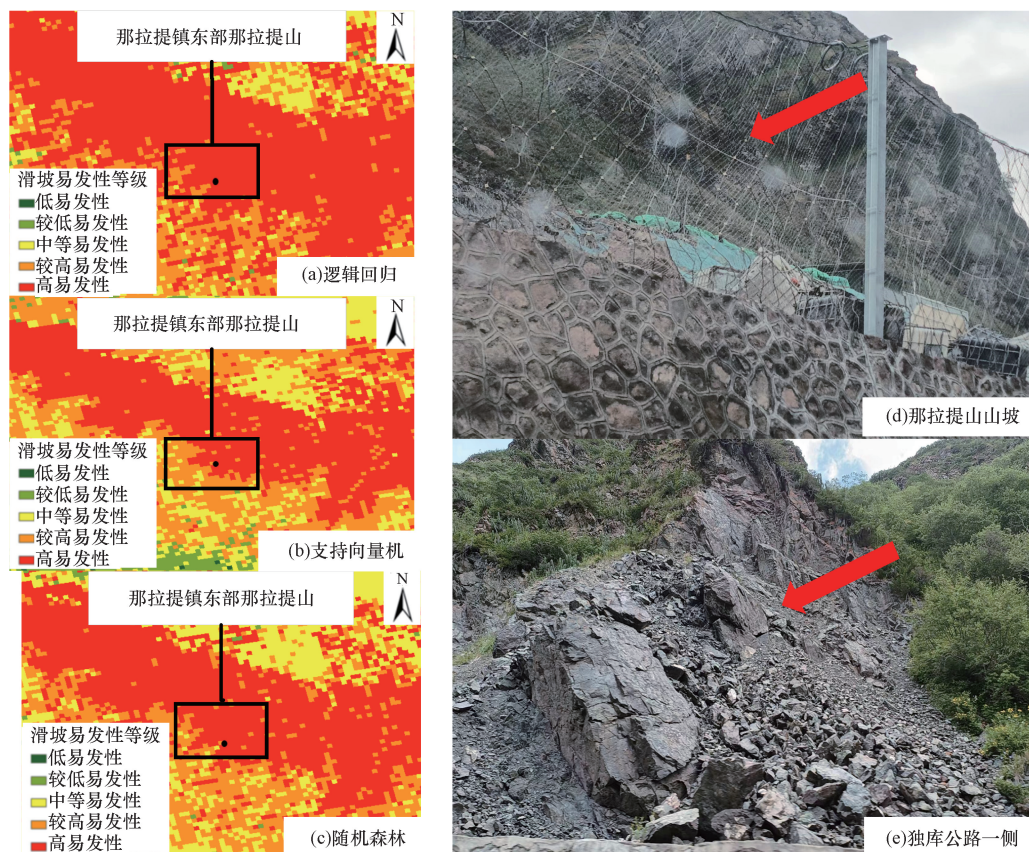


图 9 局部滑坡易发性结果与野外实地考察图

Fig. 9 Local landslide hazard results and field investigation map

地游牧居民注意安全,政府在其周边设置了警示牌和宣传标语,向居民宣传滑坡灾害的易发性和应对方法。图9(e)是那拉提镇独库公路旁边的滑坡灾害点山上滚落的石头会对公路基础设施以及行驶车辆造成严重威胁。通过图9(a)~图9(c)可知,吐尔根乡阿克克尔喀村的滑坡点均位于3种滑坡易发性评估结果的高易发区,在评估3种机器学习模型的性能时,通过对比逻辑回归(LR)、支持向量机(SVM)和随机森林(RF)模型的评估结果,发现这3种模型均对滑坡的空间位置进行了比较准确的评估,并且评估结果与实际情况高度一致,这证明了这3种模型在滑坡易发性评估领域的有效性和可靠性。但是通过观察滑坡点所位于的区域可以看出图9(c)基于支持向量机模型得到的滑坡易发性等级分布图刻画较为精细,相比其他两种模型较为准确。其能够为新源县地质灾害防治工作提供科学依据和决策支持。

4 结论

选取了高程、坡度、坡向、岩性、降雨量等17个因子作为滑坡易发性的影响因子,基于30 m网格单元,利用LR、SVM、RF共3种模型进行滑坡易发性评估研究,对3种模型滑坡易发性评估结果进行比较得到以下结论。

(1)3种算法所得的结果在低易发区和高易发区的空间位置分布上均显示出较高的一致性。具体来说,低易发区主要集中分布在巩乃斯河谷的平原地区,这些区域的地质条件相对稳定,滑坡易发性较低。而高易发区则主要集中在巩乃斯河谷以北、阿吾拉勒山的中低山丘陵区域,以及南部伊什基里克山和那拉提山分水岭的两侧,以及恰甫河以南的区域。这些区域通常地形复杂、岩性脆弱,并且受到气候变化和人为活动的多重影响,因此滑坡易发性较高。

(2)利用SPSS Pro绘制了3种算法的ROC曲线,得到LR、SVM和RF的AUC值分别为0.910、0.985和0.973。总体而言,使用新源县滑坡灾害数据集构建的3种机器学习模型均取得了令人满意的效果,其AUC值均超过0.9,这表明这3种机器学习模型都适用于新源县滑坡灾害数据集。而SVM模型在各项性能指标中表现最佳。

(3)通过对滑坡比值进行统计分析,发现LR、SVM和RF模型在高易发区的分级面积占比分别为21%、14%和16%。检验结果显示,LR、SVM和RF模型的评估结果中,分别有65%、86%和79%的滑坡点落在高易发等级区域内。因此,3种模型按照

滑坡比值由高到低的顺序排列为SVM、RF、LR和信息量模型。这说明SVM模型在新源县滑坡易发性评估结果的合理性方面优于其他两种算法。

参 考 文 献

- [1] 贺鹏,童立强,郭兆成,等. GIS支持下基于层次分析法的西藏札达地区滑坡灾害易发性评价研究[J]. 科学技术与工程, 2016, 16(25): 193-200.
He Peng, Tong Liqiang, Guo Zhaocheng, et al. GIS-supported landslide susceptibility assessment in the Zada area of Tibet based on an Analytic Hierarchy Process [J]. Science Technology and Engineering, 2016, 16(25): 193-200.
- [2] 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.
Huang Faming, Hu Songyan, Yan Xueya, et al. Landslide hazard prediction modeling and identification of main controlling factors based on machine learning [J]. Bulletin of Geological Science and Technology, 2022, 41(2): 79-90.
- [3] 薛强,张茂省,李林. 基于斜坡单元与信息量法结合的宝塔区黄土滑坡易发性评估[J]. 地质通报, 2015, 34(11): 2108-2115.
Xue Qiang, Zhang Maosheng, Li Lin. Hazard assessment of loess landslides in Baota District based on slope unit and information quantity method [J]. Geological Bulletin of China, 2015, 34(11): 2108-2115.
- [4] 李宇嘉,陈宁生,杨溢,等. 汉源县中海村“8·21”雨后型滑坡特征与成因[J]. 成都理工大学学报(自然科学版), 2022, 49(2): 185-195.
Li Yujia, Chen Ningsheng, Yang Yi, et al. Characteristics and causes of the "August 21" post-rainfall landslide in Zhonghai Village, Hanyuan County [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2022, 49(2): 185-195.
- [5] 曹文庚,潘登,徐郅杰,等. 河南省滑坡灾害易发性制图研究:多种机器学习模型的对比[J/OL]. 地质科技通报, 1-11[2025-01-07]. <https://doi.org/10.19509/j.cnki.dzkq.tb20230338>.
Cao Wengeng, Pan Deng, Xu Zhijie, et al. Hazard mapping of landslide disasters in Henan Province: a comparison of multiple machine learning models [J/OL]. Bulletin of Geological Science and Technology, 1-11[2025-01-07]. <https://doi.org/10.19509/j.cnki.dzkq.tb20230338>.
- [6] 唐川,马国超. 基于地貌单元的小区域地质灾害易发性分区方法研究[J]. 地理科学, 2015, 35(1): 91-98.
Tang Chuan, Ma Guochao. A method for zoning geological hazards in small areas based on geomorphic units [J]. Scientia Geographica Sinica, 2015, 35(1): 91-98.
- [7] 林荣福,刘纪平,徐胜华,等. 随机森林赋权信息量的滑坡易发性评估方法[J]. 测绘科学, 2020, 45(12): 131-138.
Lin Rongfu, Liu Jiping, Xu Shenghua, et al. Landslide hazard assessment method based on random forest weighted information quantity [J]. Science of Surveying and Mapping, 2020, 45(12): 131-138.
- [8] 孔嘉旭,庄建琦,彭建兵等. 基于信息量和卷积神经网络的黄土高原滑坡易发性评估[J]. 地球科学, 2023, 48(5): 1711-1729.
Kong Jiaxu, Zhuang Jianqi, Peng Jianbing, et al. Landslide hazard

- assessment in the Loess Plateau based on information quantity and convolutional neural networks [J]. *Earth Science*, 2023, 48(5): 1711-1729.
- [9] 李成林, 刘严松, 赖思翰, 等. 基于BP神经网络模型的滑坡易发性评价方法[J]. *科学技术与工程*, 2023, 23(13): 5481-5492.
- Li Chenglin, Liu Yansong, Lai Sihan, et al. Landslide susceptibility assessment method based on BP neural network model [J]. *Science Technology and Engineering*, 2023, 23(13): 5481-5492.
- [10] 何万才, 赵俊三, 林伊琳, 等. 基于证据权和支持向量机模型的威信县滑坡易发性评价[J]. *科学技术与工程*, 2023, 23(15): 6350-6360.
- He Wancai, Zhao Junsan, Lin Yilin, et al. Landslide susceptibility assessment in Weixin County based on weights of evidence and support vector machine model [J]. *Science Technology and Engineering*, 2023, 23(15): 6350-6360.
- [11] 田钦, 张彪, 郭建飞, 等. 基于信息量和逻辑回归耦合模型的滑坡易发性评价[J]. *科学技术与工程*, 2020, 20(21): 8460-8468.
- Tian Qin, Zhang Biao, Guo Jianfei, et al. Landslide susceptibility assessment based on the coupling model of information quantity and logistic regression [J]. *Science Technology and Engineering*, 2020, 20(21): 8460-8468.
- [12] 王伟中, 曹小红, 汪鑫等. 基于信息量法的新疆伊宁县滑坡灾害易发性评估[J]. *新疆地质*, 2022, 40(4): 566-573.
- Wang Weizhong, Cao Xiaohong, Wang Xin, et al. Hazard assessment of landslide disasters in Yining County, Xinjiang based on information quantity method [J]. *Xinjiang Geology*, 2022, 40(4): 566-573.
- [13] 邓念东, 李宇新, 崔阳阳, 等. 基于机器学习混合模型的滑坡易发性评价[J]. *科学技术与工程*, 2022, 22(14): 5539-5547.
- Deng Niandong, Li Yuxin, Cui Yangyang, et al. Evaluation of landslide susceptibility based on a hybrid model of machine learning [J]. *Science Technology and Engineering*, 2022, 22(14): 5539-5547.
- [14] 张成文. 新疆新源县地质灾害特征与评估研究[D]. 乌鲁木齐: 新疆农业大学, 2014.
- Zhang Chengwen. Research on characteristics and assessment of geological disasters in Xinyuan County, Xinjiang [D]. Urumqi: Xinjiang Agricultural University, 2014.
- [15] 李郎平, 兰恒星, 郭长宝, 等. 基于改进频率比法的川藏铁路沿线及邻区地质灾害易发性分区评估[J]. *现代地质*, 2017, 31(5): 911.
- Li Langping, Lan Hengxing, Guo Changbao, et al. Zoning assessment of geological hazards along and adjacent to the Sichuan-Tibet Railway based on improved frequency ratio method [J]. *Geoscience*, 2017, 31(5): 911.
- [16] 李萍, 叶辉, 谈树成. 基于层次分析法的永德县地质灾害易发性评估[J]. *水土保持研究*, 2021, 28(5): 394-399.
- Li Ping, Ye Hui, Tan Shucheng. Hazard assessment of geological disasters in Yongde County based on analytic hierarchy process [J]. *Research of Soil and Water Conservation*, 2021, 28(5): 394-399.
- [17] 李利峰, 张晓虎, 邓慧琳, 等. 基于SVM-LR融合模型的滑坡灾害易发性评价——以山阳县为例[J]. *科学技术与工程*, 2020, 20(26): 10618-10625.
- Li Lifeng, Zhang Xiaohu, Deng Huilin, et al. Landslide hazard assessment based on SVM-LR fusion model: a case study in Shanyang County [J]. *Science Technology and Engineering*, 2020, 20(26): 10618-10625.
- [18] 刘帅, 朱杰勇, 杨得虎, 马博. 不同降雨工况条件下的崩滑地质灾害危险性评价[J]. *地质科技通报*, 2024, 43(2): 253-267.
- Liu Shuai, Zhu Jieryong, Yang Dehu, et al. Hazard Assessment of Landslide Geological Hazards under Different Rainfall Conditions [J]. *Bulletin of Geological Science and Technology*, 2024, 43(2): 253-267.
- [19] 李利峰, 张晓虎, 邓慧琳, 等. 基于熵指数与逻辑回归耦合模型的滑坡灾害易发性评价——以蓝田县为例[J]. *科学技术与工程*, 2020, 20(14): 5536-5543.
- Li Lifeng, Zhang Xiaohu, Deng Huilin, et al. Evaluation of landslide susceptibility based on the coupling model of entropy index and logistic regression: a case study of Lantian County [J]. *Science Technology and Engineering*, 2020, 20(14): 5536-5543.
- [20] Zehra T K, Kursat A O, Candan G. Performance comparison of landslide susceptibility maps derived from logistic regression and random forest models in the Bolaman Basin, Türkiye[J]. *Natural Hazards Review*, 2024, 25(1): 04023054.
- [21] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995.
- [22] Shang H, Su L, Chen W, et al. Spatial prediction of landslide susceptibility using logistic regression (LR), functional trees (FTs), and random subspace functional trees (RSFTs) for Pengyang County, China[J]. *Remote Sensing*, 2023, 15(20): 15204952.
- [23] Kavzoglu T, Sahin E K, Colkesen I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression [J]. *Landslides*, 2014, 11(3): 425-439.
- [24] 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 quantity random forest model: taking the Songpan Jiaochang seismic zone as an example [J]. *Science Technology and Engineering*, 2024, 24(1): 143-154.
- [25] 吴孝情, 赖成光, 陈晓宏, 等. 基于随机森林权重的滑坡易发性评估: 以东江流域为例[J]. *自然灾害学报*, 2017, 26(5): 119-129.
- Wu Xiaoqing, Lai Chengguang, Chen Xiaohong, et al. Landslide hazard assessment based on random forest weights: a case study in the Dongjiang River Basin [J]. *Journal of Natural Disasters*, 2017, 26(5): 119-129.
- [26] Ullah K, Wang Y, Fang L R M. Multi-hazard susceptibility mapping based on convolutional neural networks [J]. *Geoscience Frontiers*, 2022, 13(5): 101425.
- [27] 周平华, 李英成, 白洁, 等. 新源县汇水单元滑坡灾害易发性评价[J]. *测绘科学*, 2017, 42(9): 82-87.
- Zhou Pinghua, Li Yingcheng, Bai Jie, et al. Evaluation of landslide hazard susceptibility in the catchment area of Xinyuan County [J]. *Science of Surveying and Mapping*, 2017, 42(9): 82-87.