水文地质工程地质
HYDROGEOLOGY & ENGINEERING GEOLOGY
中文核心期刊 CSCD核心期刊
中科双效期刊 中国科技核心期刊
Caj-cd规范获奖期刊

考虑影响因子重要性选择和土壤含水率的滑坡易发性评价

王中羽，李素敏，袁利伟，乐伟鹏

Landslide susceptibility evaluation considering the importance selection of influencing factors and soil moisture content

WANG Zhongyu, LI Sumin, YUAN Liwei, and LE Weipeng

在线阅读 View online: https://doi.org/10.16030/j.cnki.issn.1000-3665.202309009

DOI: 10.16030/j.cnki.issn.1000-3665.202309009

王中羽, 李素敏, 袁利伟, 等. 考虑影响因子重要性选择和土壤含水率的滑坡易发性评价 [J]. 水文地质工程地质, 2025, 52(3): 211-221.

WANG Zhongyu, LI Sumin, YUAN Liwei, et al. Landslide susceptibility evaluation considering the importance selection of influencing factors and soil moisture content[J]. Hydrogeology & Engineering Geology, 2025, 52(3): 211-221.

考虑影响因子重要性选择和土壤含水率的滑坡易发性评价

王中羽[1,2], 李素敏[1,3], 袁利伟[4], 乐伟鹏[1]

(1. 昆明理工大学国土资源工程学院, 云南 昆明 650093; 2. 云南省核工业二〇九地质大队, 云南 昆明 650032; 3. 云南省高校高原山区空间信息测绘技术应用工程研究中心, 云南 昆明 650093; 4. 昆明理工大学公共安全与应急管理学院, 云南 昆明 650093)

摘要: 在滑坡易发性评价体系中, 尚未形成统一的和科学的筛选影响滑坡发育因子的标准, 导致滑坡易发性评价结果的不一致性。为提高滑坡易发性评估体系的准确性, 提出一种基于机器学习的考虑因子重要性选择与土壤含水率的滑坡易发性评价体系。以云南省富民县为例, 结合遥感数据、辅助数据和现场调查数据, 编制滑坡历史记录; 利用 SAR 卫星后向散射系数和从 DEM 中提取的地表粗糙度提取土壤含水率, 通过 XGBoost 回归和 Lasso 回归模型对 15 个评价因子进行重要性排序, 并对滑坡影响因子进行多重共线性评估, 筛选出最具鉴别性的滑坡影响因子; 用轻量级梯度提升机算法和随机森林模型分别在因子重要性选择前与选择后对富民县进行滑坡易发性评价。结果表明, 土壤含水率因子对滑坡发育有较大影响; 经过因子重要性选择后的滑坡易发性评价结果准确性更高; 轻量级梯度提升机算法模型在评估中表现出优越的评估性能 (*AUC*=0.91), 表明 LightGBM 模型可以较好地应用在滑坡易发性评价中。本研究着重讨论了因子重要性选择对滑坡易发性评价体系的影响, 并有效地将面状土壤含水率因子纳入滑坡影响因子中, 提高了易发性评价结果的精确性和可靠性, 为预防滑坡灾害提供新思路。

关键词: 机器学习; 轻量级梯度提升机算法; 土壤含水率; 滑坡易发性; 地质灾害

中图分类号: P642.22 **文献标志码:** A **文章编号:** 1000-3665(2025)03-0211-11

Landslide susceptibility evaluation considering the importance selection of influencing factors and soil moisture content

WANG Zhongyu[1,2], LI Sumin[1,3], YUAN Liwei[4], LE Weipeng[1]

(1. *Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming, Yunnan* 650093, *China*; 2. *No.* 209 *Geological Team of Yunnan Nuclear Industry, Kunming, Yunnan* 650032, *China*; 3. *Yunnan Province University Plateau Mountain Area Spatial Information Surveying and Mapping Technology Application Engineering Research Center, Kunming, Yunnan* 650093, *China*; 4. *School of Public Safety and Emergency Management, Kunming University of Science and Technology, Kunming, Yunnan* 650093, *China*)

Abstract: The current landslide susceptibility assessment system lacks unified and scientifically grounded

收稿日期: 2023-09-04; **修订日期:** 2023-12-22 **投稿网址:** www.swdzgcdz.com

基金项目: 国家自然科学基金项目(41961053; 41861054); 云南省重点研发计划项目(202003AC100002)

第一作者: 王中羽(1999—), 女, 硕士研究生, 主要从事 InSAR 和地质灾害识别与监测研究。E-mail: wangzhongyu@stu.kust.edu.cn

通讯作者: 李素敏(1963—), 女, 副教授, 硕士生导师, 主要研究方向为高原山区 InSAR 数据处理及应用。E-mail: lism@kust.edu.cn

standards for selecting factors that influence landslide development, leading to inconsistencies in evaluation results. To improve the accuracy of the landslide susceptibility assessment system, a method based on machine learning (ML) , which innovatively integrates factor importance selection and soil moisture content was proposed. Focusing on landslides in the Fumin County in Yunnan Province, this study integrated remote sensing data, auxiliary data, and field survey data to compile comprehensive landslide history records. Soil moisture content factors were extracted using SAR satellite backscatter coefficients and DEM data, and 15 evaluation factors were ranked for importance using XGBoost and Lasso regression models. Multicollinearity assessments of landslide influencing factors (LIFs) were conducted to identify the most distinctive factors. The landslide susceptibility in the Fumin County was evaluated using LightGBM and Random Forest models, before and after factor importance selection. The results show a significant improvement in assessment accuracy after factor selection, with the LightGBM model exhibiting superior evaluation performance with an *AUC* value of 0.91, demonstrating its effective application in landslide susceptibility assessment. This study highlights the impact of factor importance selection on the landslide susceptibility assessment system, effectively incorporates areal soil moisture content factors into landslide influencing factors, and optimizes the precision and reliability of susceptibility assessment results, providing new insights for landslide disaster prevention.

Keywords: machine learning; LightGBM; soil moisture content; landslide susceptibility assessment; geological hazards

山体滑坡是一种常见的地质灾害。我国西南地区地质构造复杂、降雨频繁,滑坡发生频率高,给人民的生命和财产安全带来严重威胁[1－2]。滑坡易发性制图(landslide susceptibility mapping, LSM)可以有效地评估地质灾害风险,通过分析已发生滑坡区域地形地貌和水文地质特征,预测其他区域发生滑坡的概率[3]。

滑坡发育影响因子的选取是 LSM 的关键步骤。土壤含水率是影响滑坡发育的重要因素之一[4－5]。尽管 SAR 卫星数据与高精度 DEM 数据在滑坡易发性评价中已得到一定应用,但二者结合用于提取土壤含水率栅格数据,并作为滑坡易发性关键评价因子的研究还比较少。

对于滑坡影响因子的筛选,统计学模型已经取得了一定的成效。例如,Bui 等[6] 探索使用 Relief 算法对越南谅山市的滑坡影响因子进行重要性排序并删除了 3 个不重要的因子,对比因子删除前后差异,结果表明经过因子筛选后的结果更加准确;Lee 等[7] 通过卡方法筛选重要的滑坡影响因子,利用地理信息系统和遥感技术评估马来西亚槟城的滑坡风险,得出逻辑回归模型在预测滑坡风险方面优于概率模型。传统统计学模型在揭示因变量与自变量间复杂非线性关系以及处理海量数据方面存在明显局限,而机器学习(machine learning, ML)算法恰在这些方面展现出显著优势[8－9],使得 ML 在滑坡易发性评价研究中的应用价值日益凸显[10－11]。高度相关的变量组成的变量集会增加机器学习模型的复杂性,并降低可泛化性,科学地筛选影响滑坡发育因子是能否准确评价滑坡易发性的先决条件。为了提高 LSM 模型的准确性,选取 XGBoost 回归(XGBoost regression, XR)和 Lasso 回归(Lasso regression,LR)两种 ML 方法的因子选择结果进行对比研究。

由于滑坡影响因子数据的适用性、区域地质条件的复杂性、训练数据集的差异性以及算法参数设置的不同,ML 在对不同研究区域中 LSM 的表现,呈现出显著的波动性[12]。因此,对于同一地区的 LSM,采用轻量级梯度提升机算法(light gradient boosting machine, LightGBM)和随机森林模型(random forest , RF)两种不同的 ML 方法,以增加结果的准确性和科学性。

以云南省昆明市富民县为例,综合考虑影响滑坡发生的孕灾条件和诱发因素,将包括由新方法提取出的土壤含水率在内的 15 种评价因子进行相关性分析,用 XGBoost 回归和 Lasso 回归模型对因子的重要性进行排序和筛选。将因子筛选前后的结果对比,证明 ML 因子筛选的重要性与可靠性。此外,通过对比 LightGBM 和 RF 模型的评价结果,增加了结果的普适性。通过筛选后的评价因子与稳健的 ML 模型相结合,优化滑坡易发性评估体系的准确性和可靠性,为预防滑坡灾害提供新思路。

1　数据及评价技术路线

1.1　研究区概况

富民县地处昆明市西北部, 属北亚热带季风气候半湿润区。县国土面积 1 059.65 km^2。全县降雨在时空分布上不均匀, 5—10 月为雨季, 降雨量占全年降雨量的 80% ~ 88%, 2008—2021 年年降雨量在 700 ~ 900 mm 之间。由于地形复杂, 海拔差异大, “立体气候”较强, 气象灾害较为频繁。

富民县整体呈现盆岭相间的构造地貌格局。在长期的地质演化过程中, 受多期构造运动影响, 形成了以断裂构造为主导的构造体系, 主要发育南北向、北东向和北西向 3 组断裂系统。新构造运动表现为间歇性抬升特征, 造就了境内海拔高差显著、河谷深切的地形特征, 河流网络呈放射状展布, 形成了构造侵蚀地貌、构造-剥蚀地貌、构造-溶蚀地貌、侵蚀-溶蚀复合地貌以及盆地冲积地貌 5 大类地貌单元。受自然与人文因素影响, 研究区内地质灾害种类众多, 各类地质灾害分布范围广泛且发生频繁[13]。由于特殊的地质环境与较大的降雨量, 县内滑坡灾害的规模虽然不大, 但具有隐蔽性和突发性, 全县因滑坡灾害而搬迁的村落高达数十个, 直接或间接破坏公路水渠等, 经济损失高达数百万元, 给社会带来极大危害[14]。

1.2　滑坡特征

通过实地调查、历史编录数据以及 SBAS-InSAR 解译, 得到研究区的滑坡编录数据, 如图 1 所示。

研究区共发育 202 处滑坡, 以小型滑坡为主(占比 83%), 其中最大滑坡面积 13 755 m^2, 最小为 173 m^2; 滑坡形态多呈舌形, 以牵引式为主, 厚度为 0.5 ~ 11.0 m, 均属浅层滑坡。

1.3　数据来源

滑坡易发性影响因子数据源如表 1 所示, 主要包括: (1)云南省地质局 1∶5 万地质图; (2)第三次全国土地调查数据库 5 m 分辨率的 DEM 数据、道路与河流矢量数据和土地利用矢量数据; (3)富民县气象站降雨等数据; (4)昆明市地震局地震动峰值加速度数据(peak ground acceleration, PGA)。

1.4　评价技术路线

图 2 为总体方法的工作流程图, 包括不同滑坡影响因子的制备、土壤含水率的提取、多重共线性分析和多种 ML 模型下的因子重要性选择、因子重要性筛选结果对比等。

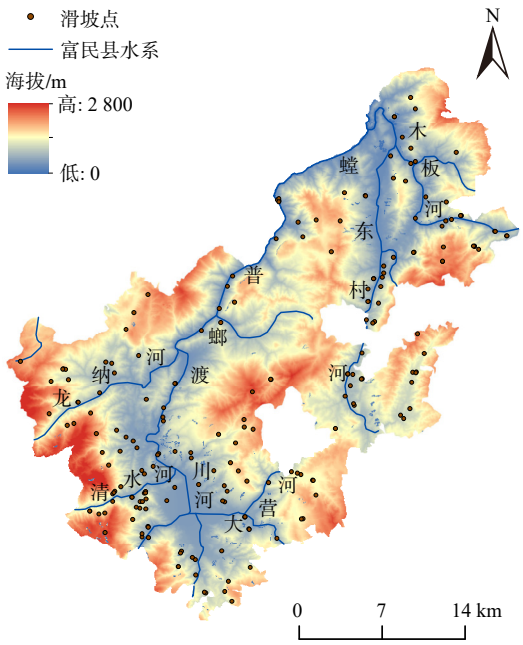


图 1　地形及地质灾害分布图

Fig. 1　Topographic and geological hazard distribution map

表 1　利用遥感和辅助数据集制备的不同的滑坡影响因子

Table 1　Different landslide impact factors prepared using remote sensing and auxiliary datasets

序号	提取因子	数据源和分辨率
1	坡度	第三次全国土地调查数据库的DEM数据(5 m×5 m)
2	坡向	
3	DEM	
4	地面起伏度	
5	平面曲率	
6	剖面曲率	
7	地层岩组	云南省地质局地质图(1∶50 000)
8	距断层距离	
9	归一化植被指数	哨兵-1A卫星数据(10 m×10 m)
10	土壤含水率	
11	土地利用	第三次全国土地调查数据库矢量数据(5 m×5 m)
12	距河流距离	
13	距道路距离	
14	GPA	昆明市地震局(2021年)
15	降雨量	富民县气象站(2022年)

2　方法

2.1　土壤含水率因子提取

土壤含水率是地表水资源形成、转换和消耗时的重要影响因素。土壤含水率过大会诱发大量的崩塌、滑坡, 也会为泥石流的发生储存大量的松散堆积体[15]。SAR 卫星后向散射系数主要由土壤含水率、地表粗糙度、植被覆盖度主导[16]。由于高原地区的植被对 SAR 土壤含水率反演的影响较小[17], 因此可以忽略植被覆盖

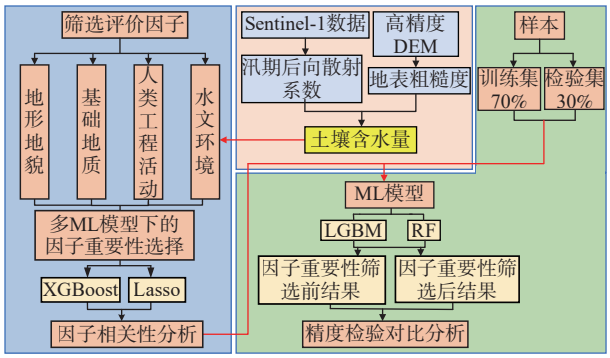


图 2 滑坡易发性评价集成框架流程

Fig. 2 Integrated framework process for landslide susceptibility assessment

对数值的影响。将欧洲航天局(European Space Agency, ESA)的 sentinel-1A 卫星数据用 ENVI 与 ArcGIS 软件计算处理得到土壤含水量因子,计算式如下:

$$\rho' = \rho - \rho^{\circ} \tag{1}$$

式中:ρ——汛期时序范围内后向散射系数;

ρ°——5 m 高精度 DEM 提取的地表粗糙度;

ρ'——土壤水含量。

将剔除地表粗糙度的后向散射系数视为土壤水含量,得到土壤含水率影响因子。

2.2 滑坡评价因子重要性选择及相关性检验

在现阶段的滑坡易发性模拟实践中,对于影响滑坡发育的变量选取没有统一的标准。筛选重要的滑坡易发性评价因子是获得可靠评价结果的重要步骤之一。由高度相关的因子组成的变量集有可能增加 ML 模型的复杂性并降低准确性。故需要对因子进行重要性选择和相关性检验。

2.2.1 因子重要性选择

因子重要性选择是通过优化变量集合来提升模型预测精度并降低复杂度的关键过程,因此选用 XGBoost 回归和 Lasso 回归模型来评价各影响因子在滑坡发育中的重要性。

(1)XGBoost 回归

XR 是一种基于决策树的集成式梯度提升算法。其核心在于通过增强集成技术将多个弱学习器结合,通过多次迭代生成一个强学习器。计算式如下:

$$L^{(k)} = \sum_{i=1}^{n} l\left(XR_i, \widehat{XR}_i^{(k-1)} + f_k(x_i)\right) + \Omega(f_k) \tag{2}$$

$$\widehat{XR}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \theta \tag{3}$$

其中,$L^{(k)}$是目标函数,包括损失函数和正则项;$\sum_{i=1}^{n} l\left(XR_i, \widehat{XR}_i^{(k-1)} + f_k(x_i)\right)$是损失函数,表示$n$个样本经过$k$次迭代后第$i$个样本的预测值;$f_k(x_i)$表示$k$棵树对第$i$个样本进行预测;$\Omega(f_k)$为正则项,表示第$k$次迭代的正则项,用来控制模型的复杂度;通过上述多次迭代最小化目标函数,最终输出模型的预测值$\widehat{XR}_i$;其中f_k表示第k棵回归树,k为回归树的数量,θ表示所有可能回归树的空间。

(2)Lasso 回归

最小绝对值收缩和选择算子。Lasso 在损失函数的基础上,通过引入惩罚函数$L1$范数项约束模型的复杂程度,使得不显著的回归系数被压缩为 0,同时也可以减少对重要回归系数的过度压缩,从而最终获得一个变量较少且更有效率的模型[18]。计算式如下:

$$LR(\boldsymbol{w}) = \sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{n} w_j x_{ij}\right)^2 + \alpha \sum_{j=1}^{w} |w_j| \tag{4}$$

式中:m——自变量的数量;

n——特征数量;

$\boldsymbol{w}$——回归系数向量;

x——特征矩阵;

y_i——样本的真实值;

α——控制回归系数的压缩程度。

2.2.2 因子相关性检验

为消除变量间多重共线性对模型评估精度的影响,采用皮尔逊相关系数法对 15 个滑坡影响因子进行共线性检验。

2.3 滑坡易发性评价模型

由于滑坡的形成与多种因素有关,各种因素之间存在复杂的交互效应。LightGBM 模型可以通过构建复杂的树结构来捕捉这些交互,该模型 2017 年才刚被提出,少有学者使用。而 RF 模型在之前学者有关 LSM 的研究中表现出其优越的性能,故选取 LightGBM 和 RF 模型作为滑坡易发性评价模型。为了提高模型的准确性,使用网格搜索算法对各个模型中的参数进行调整优化。

(1)LightGBM 模型

LightGBM 模型是梯度提升决策树(gradient boosting decision tree, GBDT)算法框架的高效改进实现,其在传统 GBDT 算法的基础上创新性地融入了基于梯度的单侧采样、互斥特征捆绑和特征直方图等优化方法,显著提升了算法的性能和效率[19]。优化使 LightGBM 算法具有低内存消耗和高训练效率,更适合处理滑坡

易发性评价这类大规模数据。通过网格搜索算法, 得出当 LightGBM 模型中的弱学习机器个数为 200, 树的最大深度为 8, 迭代中随机选择的参数比例为 0.5 时模型精度最高, 达到 88.97%。

(2)随机森林模型

随机森林是一种基于决策树的并行集成学习算法, 其核心在于通过构建多棵决策树形成森林结构, 并在节点划分时引入随机属性选择机制[20]。通过网络搜索算法优化, 当 RF 模型的弱学习的机器个数为 10, 决策树的最大深度为 2, 最大特征数为 5, 叶子节点最少样本数为 1, 内部节点再划分所需最小样本数为 2 时模型精度最高, 达到 86.23%。

2.4　模型性能评价方法

ROC 曲线的 *AUC* 值(式 5)具有不受临界约束和试验准确性高的优点, 可用于定量评估模型的准确性, 是 LSM 中最常用的性能指标。*AUC* 的取值范围为 0.5 ~ 1.0, 当 $AUC \geqslant 0.7$ 为可接受, $AUC \geqslant 0.8$ 为宜, $AUC \geqslant 0.9$ 时为极好。

$$AUC = \frac{\sum TP + \sum TN}{TP + FP + TN + FN} \tag{5}$$

式中: *TP*——正确分类的滑坡单元;

TN——正确分类的非滑坡单元;

FP——滑坡单元被错误地分类为非滑坡;

FN——非滑坡单元被错误地分类为滑坡。

3　结果

结合研究区域面积及原数据精度, 以 20 m×20 m 的栅格单元为最小评价单元, 将 202 个滑坡点与同等数量的非滑坡点转换为栅格数据, 共计 1 108 个栅格单元, 其中 70% 的样本用于模型训练, 剩下的 30% 用于测试模型的训练精度。并根据以前研究人员的经验和滑坡点分布模式对每个滑坡影响因子进行重分类, 再对所有图层的子类别进行赋值处理, 如图 3 所示。

3.1　滑坡评价因子重要性选择及相关性检验结果

通过野外调查研究及已有相关调查报告, 综合考虑研究区的地形地貌、基础地质、水文条件等因素, 最终选取 15 个滑坡影响因子: 剖面曲率(A1)、坡度(A2)、平面曲率(A3)、坡向(A4)、土壤含水率(A5)、土地利用方式(A6)、归一化植被指数(A7)、降水(A8)、距河流距离(A9)、地震动峰值加速度(A10)、距断层距离(A11)、地面起伏度(A12)、岩性(A13)、高程(A14)、距道路距离(A15), 15 个影响因子分别用 A1—A15 表示, 以便处理。

(1) 因子重要性选择结果

图 4 展示了 XR 和 LR 方法评估的各影响因子相对重要性, 重点在于纵向比较不同因子选择方法的数值差异, 而非横向对比方法间的数值大小, 以揭示各因子对滑坡发育的潜在贡献。结果表明, 距河流距离、归一化植被指数和土壤含水率被判定为关键影响因子。同时, 因子选择结果显示, 剖面曲率(A1)、平面曲率(A3)和地面起伏度(A12)贡献度最低, 将被剔除。

(2) 因子相关性检验结果

图 5 为影响因子的相关性热力图, 根据相关性检验结果, 地面起伏度(A12)与坡度(A2)有很高的相关性, 相关系数高达 0.83。再结合前文的因子重要性选择结果, 将地面起伏度(A12)、平面曲率(A3)、剖面曲率(A1)从指标体系中删除。剔除 3 个影响因子后, 对剩余的因子再次进行相关性检验, 如图 5(b)所示。所提出的土壤含水率因子 A5 与各因子间的相关系数均小于 0.3。结果表明各变量之间具有相互独立性。因此, 可以满足模型的需求。

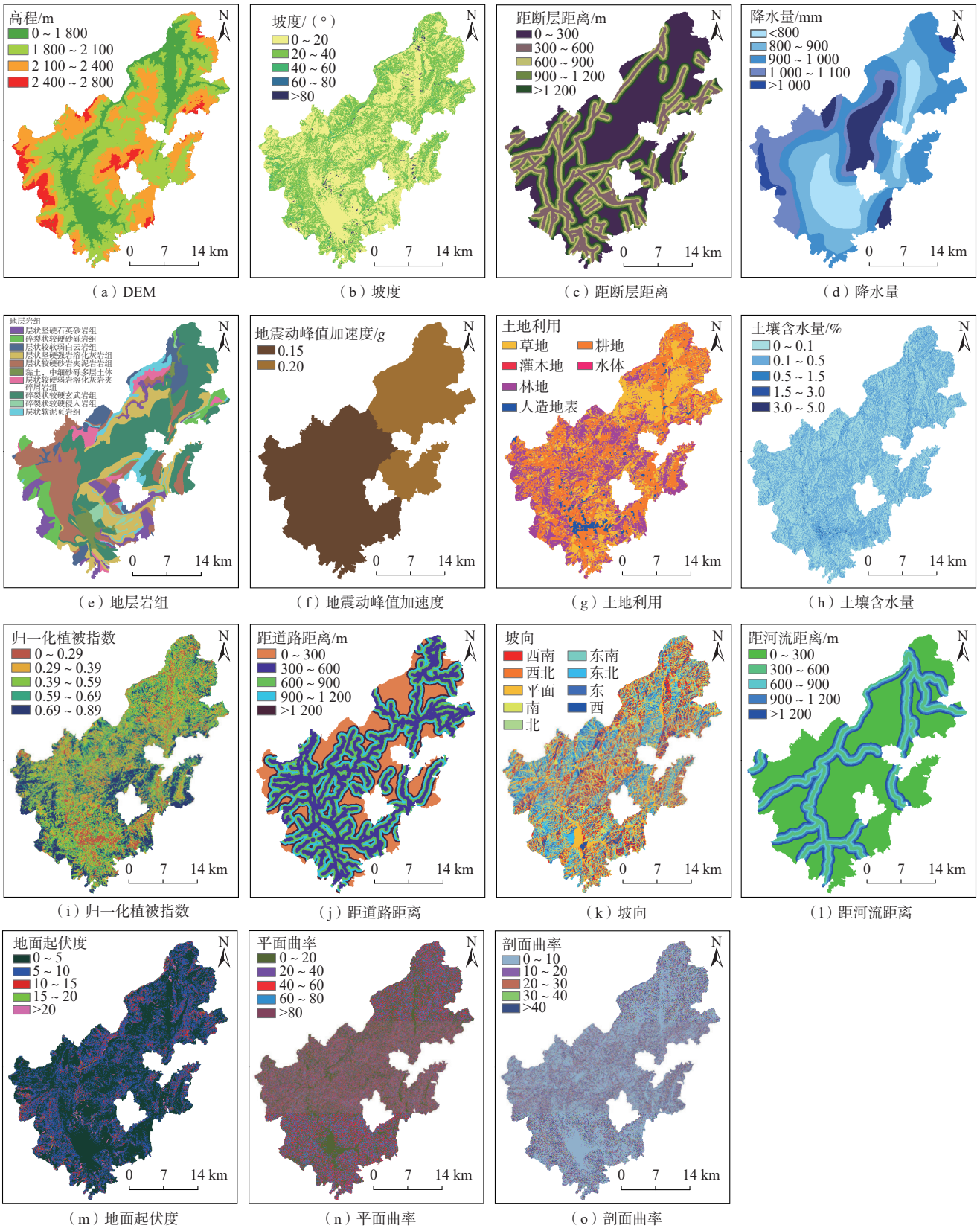
3.2　模型性能评价

将未参与训练的 30% 滑坡与非滑坡的栅格单元作为检验样本, 共计 332 个; 将参与训练的 70% 滑坡与非滑坡的栅格单元作为成功样本, 共计 776 个。研究成功率曲线如图 6(a)所示, 预测率曲线如图 6(b)所示。RF 和 LightGBM 模型的成功率曲线 *AUC* 值分别为 0.89 和 0.94。预测率曲线 *AUC* 值分别为 0.86 和 0.91。成功率与预测率的 *AUC* 最小值为 0.86, 表明 2 种模型的评价结果与实际情况的高度一致性, 准确地评价了富民县的滑坡易发性。其中 LightGBM 模型的评价精度最高, 表明 LightGBM 模型在 LSM 领域具有良好的适用性和应用潜力。

3.3　滑坡易发性制图

由 LightGBM 和 RF 模型评价的因子重要性筛选前后的滑坡易发性对比如图 7 所示, 在 GIS 环境中, 各个模型的概率值用自然断点法分为 5 类: 极低易发区、低易发区、中易发区、高易发区和极高易发区。高易发区和极高易发区主要集中分布于研究区西南部和东北部, 均位于地形陡峭的高原山地地带。研究区中部和东南部地区地势相对平缓, 且年平均降雨量相对较小, 主要为极低易发区和低易发区。

利用 GIS 多值提取分别将包含 15 个影响因子与包含 12 个影响因子的滑坡敏感性区域信息提取至滑坡点, 如表 2 所示。



（a）DEM　（b）坡度　（c）距断层距离　（d）降水量

（e）地层岩组　（f）地震动峰值加速度　（g）土地利用　（h）土壤含水量

（i）归一化植被指数　（j）距道路距离　（k）坡向　（l）距河流距离

（m）地面起伏度　（n）平面曲率　（o）剖面曲率

图 3　15 个滑坡影响因子专题图

Fig. 3　15 thematic maps of landslide impact factors

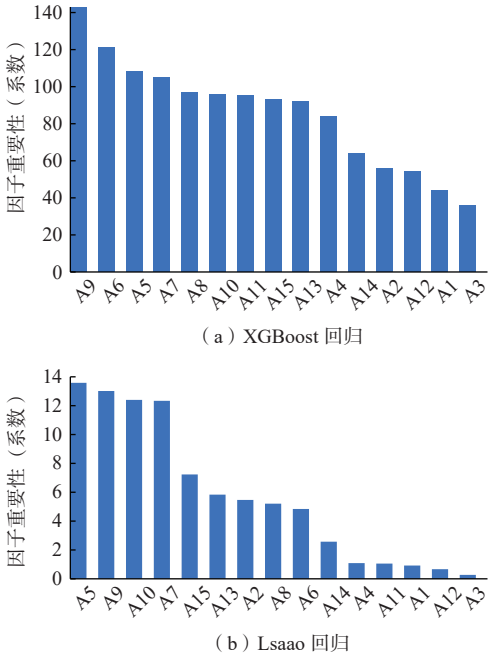


图 4　因子重要性选择结果图

Fig. 4　Results of factor importance selection

从滑坡点的空间分布特征来看, 各模型在极高易发区和高易发区识别的滑坡点数量显著多于中易发区、低易发区和极低易发区。LightGBM 模型的高易发区和极高易发区中的滑坡点占比最多, 高达 80%。RF 模型的评价结果显示, 随着易发程度等级的提升, 滑坡频率比值呈现显著递增趋势, 其中极低、低和中易发区的频率比值均小于 1。包含 15 个影响因子和包含 12 个影响因子的 LightGBM 模型的极高易发区的频率比值分别为 2.65 和 3.80 均高于 RF 模型在极高易发区的频率比值(1.86 和 3.27)。结果表明, 经过重要性筛选后的各模型的频率比值均有明显的提升, 且 LightGBM 模型的频率比值整体优于 RF 模型。

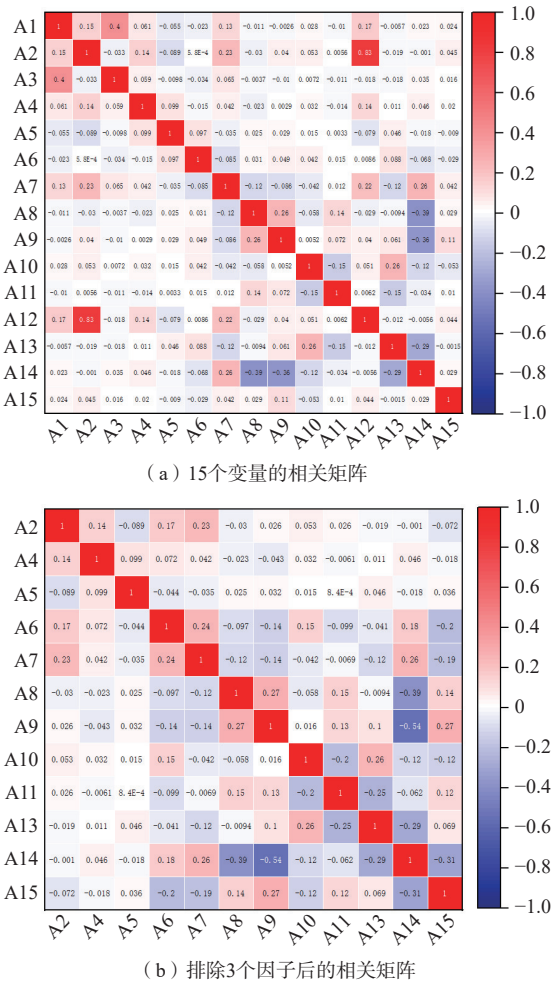


图 5　相关系数矩阵

Fig. 5　Correlation coefficient matrix

从细节来看, 图 8 是在 LightGBM 模型下因子选择前后的两个重点区域的滑坡易发性差异对比图。从图 8(c)可以看出 A1 和 A2 滑坡点位于普渡河旁的公路西岸, 为东向坡, 坡体均呈舌状, 坡度为 20°。此区域整体属于块状、碎裂状较硬玄武岩组, 节理裂隙

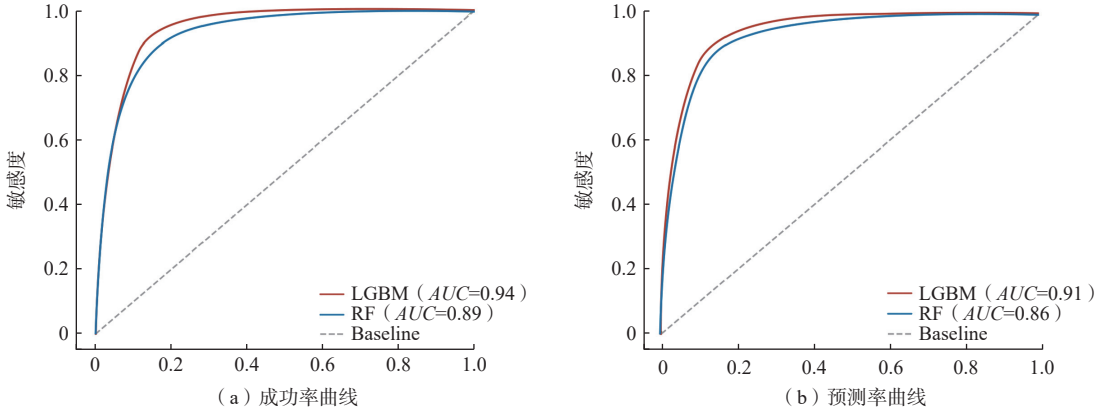
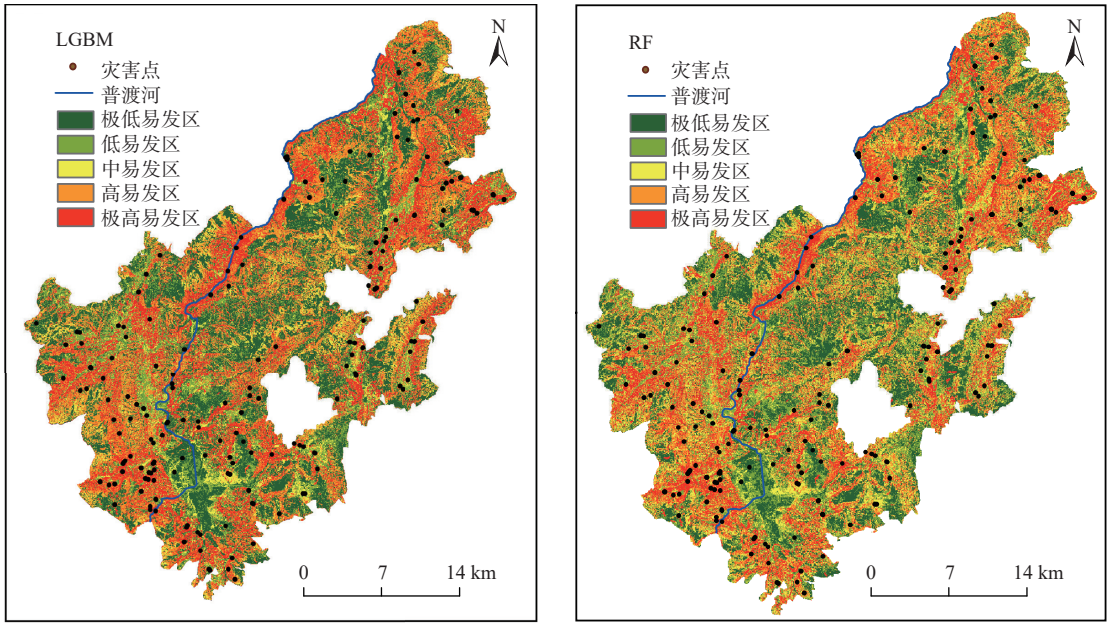
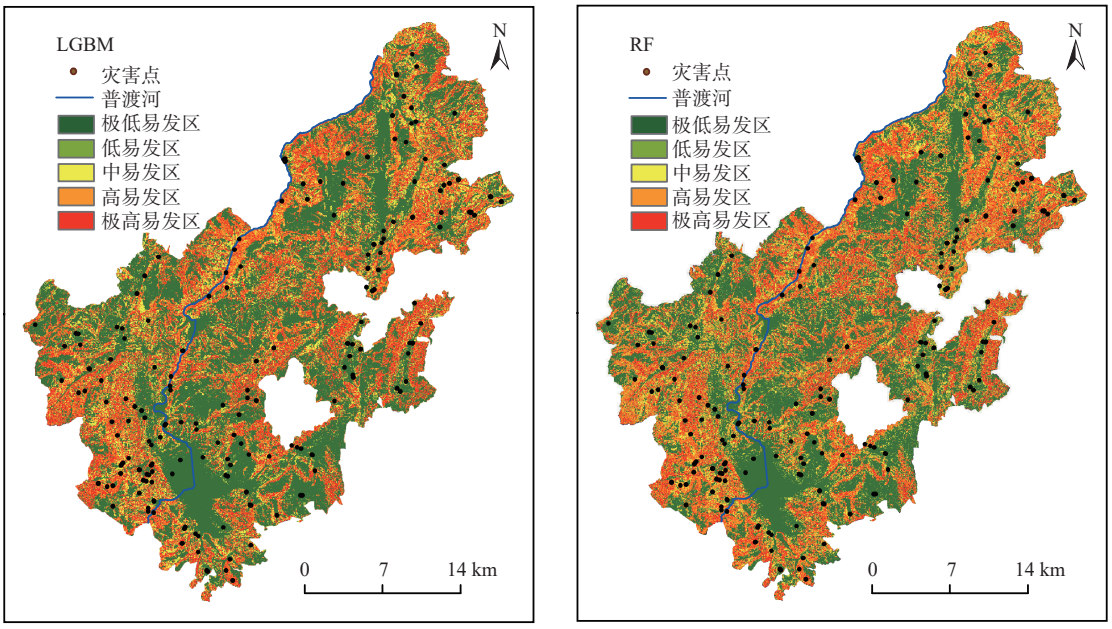


图 6　ROC 曲线

Fig. 6　ROC curve



（a）15个影响因子参与的评价结果



（b）12个影响因子参与的评价结果

图 7　因子重要性筛选前后的评价结果对比图

Fig. 7　Comparison of evaluation results before and after factor importance screening

发育。图 8（a）为因子重要性选择前的区域滑坡易发性图，图中显示，两滑坡点均落于低易发性区域中，对整个公路及周边区域均缺乏预测。图 8（f）的滑坡属于块状、碎裂状较硬玄武岩组，滑坡周围多居民与农田；据现场调查，此地小型滑坡频发，已对当地居民生活造成较大影响；由因子重要性选择前后的结果图 8（d）和图 8（e）对比来看，因子重要性选择后的易发性评价图更加精确。

总体来看，经过重要性筛选后的因子的模型优于其包含全部影响因子的模型。

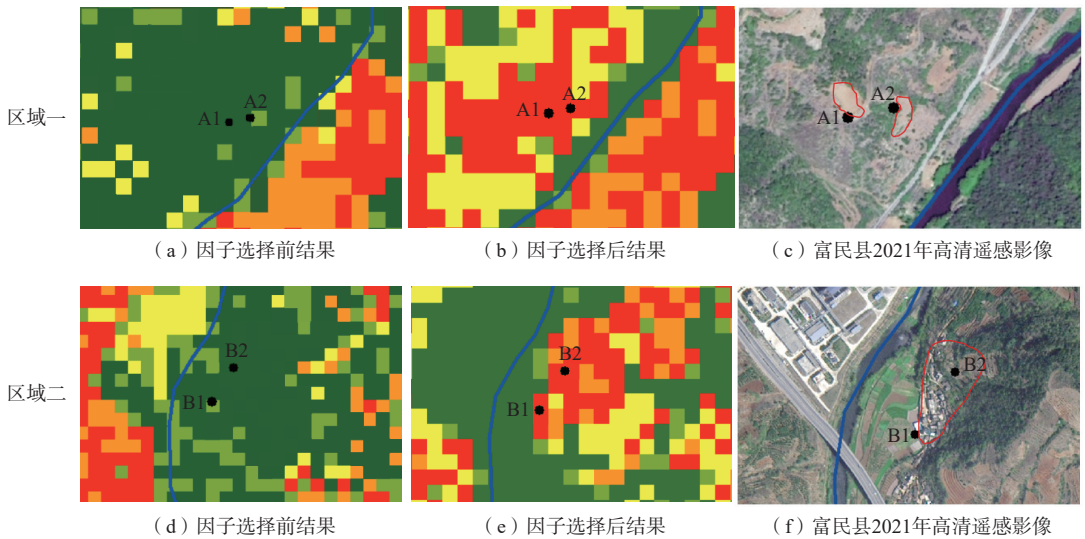
4　讨论

滑坡易发性评价的结果受滑坡影响因子的直接影响。在两个因子重要性选择模型的结果中，剖面曲率、平面曲率、地形起伏度均排名靠后。且在影响因子相关性检验中地形起伏度和坡度有较大的相关性。在本研究中，距河流距离、土壤含水率、归一化植被指数是因子特征选择排名前 3 位的滑坡影响因子。

表 2　因子重要性筛选前后的富民县滑坡易发性评价频率比

Table 2　Frequency ratio of landslide susceptibility evaluations in the Fumin County before and after factor importance screening

15个影响因子评价结果							12个影响因子评价结果						
模型	分级	面积/km²	容积率/%	滑坡点/个	滑坡比率/%	频率比	模型	分级	面积/km²	容积率/%	滑坡点/个	滑坡比率/%	频率比
LightGBM	极低	263.433	26.47	14	6.93	0.26	LightGBM	极低	300.530	30.20	15	7.43	0.25
	低	258.143	25.95	33	16.34	0.63		低	297.597	29.91	13	6.44	0.21
	中等	123.512	12.42	30	14.85	1.20		中等	111.805	11.24	16	7.92	0.70
	高	174.785	17.57	31	15.35	0.87		高	132.138	13.28	40	19.80	1.49
	极高	174.982	17.59	94	46.53	2.65		极高	152.785	15.36	118	58.41	3.80
RF	极低	132.207	13.29	6	2.97	0.22	RF	极低	262.435	26.38	19	9.41	0.36
	低	227.083	22.83	46	22.77	1.00		低	240.375	24.16	14	6.93	0.28
	中等	204.060	20.51	27	13.37	0.65		中等	147.362	14.81	12	5.94	0.40
	高	250.513	25.18	58	28.71	1.14		高	189.765	19.07	54	26.73	1.40
	极高	180.992	18.19	65	32.18	1.77		极高	154.918	15.57	103	50.99	3.27



（a）因子选择前结果　（b）因子选择后结果　（c）富民县2021年高清遥感影像

（d）因子选择前结果　（e）因子选择后结果　（f）富民县2021年高清遥感影像

图 8　LightGBM 模型下的因子选择前后重点区域滑坡易发性差异对比图

Fig. 8　Comparison of landslide susceptibility differences in key areas before and after factor selection under LightGBM model

土壤含水率对滑坡易发性的预测具有重要意义。强降雨、湿润的前期土壤水分条件和地下水位的深度通常共同导致滑坡的发生。土壤含水率是影响斜坡稳定性的关键因素，它通过降低岩土体强度和增加剪应力导致斜坡失稳。滑坡影响因子重要性选择结果显示，土壤含水率为诱发滑坡的重要影响因素。

约 80% 的历史滑坡点分布于中至极高易发区，与预测结果呈现显著空间一致性。其余约 20% 的历史滑坡点处于极低和低易发区域，可能是由于现场核实的滑坡清单中的一些滑坡没有覆盖足够的空间范围，无法由更粗分辨率的滑坡影响因子准确预测，导致了训练数据和影响因子的不确定性，使模型在学习历史滑坡和滑坡影响因子之间复杂非线性关系时受到阻碍，从而影响到区域尺度空间预测的准确性。

仅用一年的归一化植被指数和土地利用数据，对易发性模型的样本训练与结果输出，可能会影响预测结果的准确性。多时相滑坡影响因子数据能否可以对评价结果有所改进，是一个值得深究的问题。

仅用 *AUC* 值评价模型的性能。在以后的研究中，可以考虑基于迭代的交叉验证的方法来评估评价模型的鲁棒性，也可以提供评价模型的稳健性与准确性的进一步信息。

5　结论

（1）通过结合汛期 SAR 卫星后向散射系数和由 DEM 提取的地表粗糙度，提取了区域内的面状土壤含水率因子。在 XGBoost 回归和 Lasso 回归两种机器学习方法的重要性排序下，土壤含水率的重要性均位于前三，结果表明，该方法提取的土壤含水率因子在滑坡易发性评价中具有重要应用价值。

(2)用 2 种不同的 ML 模型对滑坡因子重要性进行排序,对比因子重要性筛选前后的试验结果,明确了因子筛选对于提高评估可靠性的重要作用。

(3)利用网格搜索算法优化的 LightGBM 和 RF 模型来评估富民县的滑坡易发性。LightGBM 模型(*AUC*=0.91)的评估结果高于 RF 模型(*AUC*=0.86)的评估结果。说明 LightGBM 模型在滑坡易发性评价中具有较强的应用潜力,有助于土地利用规划、防震减灾等活动。

参考文献(References):

[1] 陈涛,钟子颖,牛瑞卿,等. 利用深度信念网络进行滑坡易发性评价[J]. 武汉大学学报(信息科学版),2020, 45(11):1809 − 1817. [CHEN Tao, ZHONG Ziying, NIU Ruiqing, et al. Mapping landslide susceptibility based on deep belief network[J]. Geomatics and Information Science of Wuhan University, 2020, 45(11):1809 − 1817. (in Chinese with English abstract)]

[2] 张玘恺,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 landslide susceptibility mapping rapid assessment models in Jiuzhaigou County, Sichuan province, China[J]. Chinese Journal of Rock Mechanics and Engineering, 2020, 39(8):1595 − 1610. (in Chinese with English abstract)]

[3] 杜国梁,杨志华,袁颖,等. 基于逻辑回归–信息量的川藏交通廊道滑坡易发性评价[J]. 水文地质工程地质,2021, 48(5):102 − 111. [DU Guoliang, YANG Zhihua, YUAN Ying, et al. Landslide susceptibility mapping in the Sichuan-Tibet traffic corridor using logistic regression-information value method[J]. Hydrogeology & Engineering Geology, 2021, 48(5):102 − 111. (in Chinese with English abstract)]

[4] 唐鸿磊,陈菊,沈春颖,等. 饱和导水率异质性对黄土高原浅层滑坡的影响[J]. 清华大学学报(自然科学版),2023, 63(12):1946 − 1960. [TANG Honglei, CHEN Ju, SHEN Chunying, et al. Impacts of heterogeneity of saturated hydraulic conductivity on the shallow landslides on the Loess plateau[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(12):1946 − 1960. (in Chinese with English abstract)]

[5] 赵彬如,陈恩泽,戴强,等. 基于水文-气象阈值的区域降雨型滑坡预测研究[J]. 测绘学报,2022, 51(10):2216 − 2225. [ZHAO Binru, CHEN Enze, DAI Qiang, et al. Study on prediction of regional rainfall-induced landslides based on hydro-meteorological threshold[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(10):2216 − 2225. (in Chinese with English abstract)]

[6] BUI D T, PRADHAN B, REVHAUG I, et al. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam)[J]. Geomatics, Natural Hazards and Risk, 2015, 6(3):243 − 271.

[7] LEE S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data[J]. International Journal of Remote Sensing, 2005, 26(7):1477 − 1491.

[8] 黄发明,曹中山,姚池,等. 基于决策树和有效降雨强度的滑坡危险性预警[J]. 浙江大学学报(工学版),2021, 55(3):472 − 482. [HUANG Faming, CAO Zhongshan, YAO Chi, et al. Landslides hazard warning based on decision tree and effective rainfall intensity[J]. Journal of Zhejiang University (Engineering Science), 2021, 55(3):472 − 482. (in Chinese with English abstract)]

[9] 李铁锋,丛威青. 基于 Logistic 回归及前期有效雨量的降雨诱发型滑坡预测方法[J]. 中国地质灾害与防治学报,2006, 17(1):33 − 35. [LI Tiefeng, CONG Weiqing. A method for rainfall-induced landslides prediction based on Logistic regression and effective antecedent rainfall[J]. The Chinese Journal of Geological Hazard and Control, 2006, 17(1):33 − 35. (in Chinese with English abstract)]

[10] 李文彦,王喜乐. 频率比与信息量模型在黄土沟壑区滑坡易发性评价中的应用与比较[J]. 自然灾害学报,2020, 29(4):213 − 220. [LI Wenyan, WANG Xile. Application and comparison of frequency ratio and information value model for evaluating landslide susceptibility of loess Gully Region[J]. Journal of Natural Disasters, 2020, 29(4):213 − 220. (in Chinese with English abstract)]

[11] 齐信,黄波林,刘广宁,等. 基于 GIS 技术和频率比模型的三峡地区秭归向斜盆地滑坡敏感性评价[J]. 地质力学学报,2017, 23(1):97 − 104. [QI Xin, HUANG Bolin, LIU Guangning, et al. Landslide susceptibility assessment in the Three Gorges area, China, Zigui synclinal basin, using GIS technology and frequency ratio model[J]. Journal of Geomechanics, 2017, 23(1):97 − 104. (in Chinese with English abstract)]

［12］ 刘福臻, 王灵, 肖东升. 机器学习模型在滑坡易发性评价中的应用 [J]. 中国地质灾害与防治学报, 2021, 32(6): 98 − 106. [LIU Fuzhen, WANG Ling, XIAO Dongsheng. Application of machine learning model in landslide susceptibility evaluation[J]. The Chinese Journal of Geological Hazard and Control, 2021, 32(6): 98 − 106. (in Chinese with English abstract)]

［13］ 杨兰芳, 李跃琼, 董晓霞, 等. 富民县发展山区农业的思考——以永定街道为例 [J]. 中国农技推广, 2023, 39(1): 12 − 13. [YANG Lanfang, LI Yueqiong, DONG Xiaoxia, et al. Thoughts on developing mountain agriculture in Fumin County—taking Yongding street as an example[J]. China Agricultural Technology Extension, 2023, 39(1): 12 − 13. (in Chinese)]

［14］ 张洪, 陈凤娇. 低丘缓坡山地建设开发景观生态规划研究——以云南省富民县哨箐区块为例 [J]. 中国市场, 2016(4): 156 − 165. [ZHANG Hong, CHEN Fengjiao. Study on landscape ecological planning of construction and development in low hills and gentle slopes: A case study of shaoqing block in Fumin County, Yunnan Province[J]. China Market, 2016(4): 156 − 165. (in Chinese)]

［15］ 刘勇, 魏良帅, 黄安邦, 等. 气候变化下长江源土壤水时空演化及其环境响应 [J]. 水文地质工程地质, 2023, 50(5): 39 − 52. [LIU Yong, WEI Liangshuai, HUANG Anbang, et al. Spatial and temporal evolution of soil water and its response to the environment in the Yangtze River source area under climate change[J]. Hydrogeology & Engineering Geology, 2023, 50(5): 39 − 52. (in Chinese with English abstract)]

［16］ 陈庭轩, 李素敏, 袁利伟. InSAR 在云南德钦县一中河泥石流物源汇集及失稳模式中的探究 [J]. 大地测量与地球动力学, 2022, 42(2): 158 − 164. [CHEN Tingxuan, LI Sumin, YUAN Liwei. InSAR study on substance source collection and instability mode of Yizhong River debris flow in Deqin, Yunnan Province[J]. Journal of Geodesy and Geodynamics, 2022, 42(2): 158 − 164. (in Chinese with English abstract)]

［17］ 陈晓东, 郭金运, 孙明智, 等. 青藏高原区域不同地表类型对应后向散射系数的时变分析 [J]. 武汉大学学报(信息科学版), 2023, 48(5): 730 − 740. [CHEN Xiaodong, GUO Jinyun, SUN Mingzhi, et al. Time-varying analysis of backscatter coefficient corresponding to different surface types in the Tibetan Plateau[J]. Geomatics and Information Science of Wuhan University, 2023, 48(5): 730 − 740. (in Chinese with English abstract)]

［18］ 张钟远, 邓明国, 徐世光, 等. 镇康县滑坡易发性评价模型对比研究 [J]. 岩石力学与工程学报, 2022, 41(1): 157 − 171. [ZHANG Zhongyuan, DENG Mingguo, XU Shiguang, et al. Comparison of landslide susceptibility assessment models in Zhenkang County, Yunnan Province, China[J]. Chinese Journal of Rock Mechanics and Engineering, 2022, 41(1): 157 − 171. (in Chinese with English abstract)]

［19］ YAO Xin, THAM L G, DAI F C. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China[J]. Geomorphology, 2008, 101(4): 572 − 582.

［20］ 贾雨霏, 魏文豪, 陈稳, 等. 基于 SOM-I-SVM 耦合模型的滑坡易发性评价 [J]. 水文地质工程地质, 2023, 50(3): 125 − 137. [JIA Yufei, WEI Wenhao, CHEN Wen, et al. Landslide susceptibility assessment based on the SOM-I-SVM model[J]. Hydrogeology & Engineering Geology, 2023, 50(3): 125 − 137. (in Chinese with English abstract)]

编辑: 王支农