

不同机器学习模型在湖北巴东滑坡易发性评价中的应用

陈国圳, 刘军旗, 崔婧慧, 喻书鹏, 张小虎, 朱宁辉, 宋雨欣

Application of different machine learning models in landslide susceptibility assessment in Badong County, Hubei Province

CHEN Guozhen, LIU Junqi, CUI Jinghui, YU Shupeng, ZHANG Xiaohu, ZHU Ninghui, and SONG Yuxin

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.202310011>

您可能感兴趣的其他文章

Articles you may be interested in

机器学习模型在滑坡易发性评价中的应用

Application of machine learning model in landslide susceptibility evaluation

刘福臻, 王灵, 肖东升 中国地质灾害与防治学报. 2021, 32(6): 98-106

岷江上游汶川地震前后泥石流易发性评价

Susceptibility assessment of debris flow in the upper reaches of the Minjiang River before and after the Wenchuan earthquake

赵佳忆, 田述军, 李凯, 侯鹏鹞 中国地质灾害与防治学报. 2024, 35(1): 51-59

基于机器学习的区域滑坡危险性评价方法综述

A review of the methods of regional landslide hazard assessment based on machine learning

方然可, 刘艳辉, 黄志全 中国地质灾害与防治学报. 2021, 32(4): 1-8

独库高速公路克扎依—巩乃斯段雪崩易发性评价

Avalanche susceptibility evaluation of the Kezhayi to Gongnaisi section of the Duku expressway

程秋连, 刘杰, 杨治伟, 张天意, 王斌 中国地质灾害与防治学报. 2024, 35(1): 60-71

基于XGBoost模型的三峡库区燕山乡滑坡易发性评价与区划

Assessment of landslide susceptibility mapping based on XGBoost model: A case study of Yanshan Township

吴宏阳, 周超, 梁鑫, 袁鹏程, 余蓝冰 中国地质灾害与防治学报. 2023, 34(5): 141-152

基于滑坡破坏模式分析的易发性评价

Evaluation of landslide susceptibility based on landslide failure mode analysis: A case study of the left bank of Xietan River in the first section of Three Gorges Reservoir

朱宇航, 黄海峰, 殷坤龙, 郭子正, 郭飞, 赖鹏 中国地质灾害与防治学报. 2023, 34(2): 156-166



关注微信公众号, 获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.202310011

陈国圳, 刘军旗, 崔婧慧, 等. 不同机器学习模型在湖北巴东滑坡易发性评价中的应用[J]. 中国地质灾害与防治学报, 2025, 36(3): 140-150.

CHEN Guozhen, LIU Junqi, CUI Jinghui, et al. Application of different machine learning models in landslide susceptibility assessment in Badong County, Hubei Province[J]. The Chinese Journal of Geological Hazard and Control, 2025, 36(3): 140-150.

不同机器学习模型在湖北巴东滑坡易发性评价中的应用

陈国圳, 刘军旗, 崔婧慧, 喻书鹏, 张小虎, 朱宁辉, 宋雨欣

(中国地质大学(武汉)湖北巴东地质灾害国家野外科学观测研究站, 湖北 武汉 430074)

摘要: 中国是世界上发生滑坡灾害最频繁的国家之一, 滑坡易发性评价有助于防灾减灾工作。由于不同机器学习模型在不同区域的适配程度不同, 为更好开展湖北省巴东县的滑坡灾害防治工作, 选取坡度、坡向、曲率、起伏度、地层、覆盖层、归一化植被指数、道路密度、水系密度、斜坡结构 10 个影响因子, 采用逻辑回归、支持向量机、多层感知机和随机森林 4 种模型进行滑坡易发性评价。并通过受试者工作特征曲线、均方误差与决定系数等指标、滑坡-研究区占比 3 种评价方式用于评价模型精度。试验结果表明: 不同模型在不同评价方式中存在差异, 但总体而言, RF 模型精度最高且绘制出的易发性分区图更合理。4 个模型绘制的易发性区域分布图相似, 极高易发区和高易发区主要分布于南边沿江地区, 西南沿岸的官渡口镇、焦家湾村等附近地区表现出较高易发性, 该评价结果可以为巴东县的滑坡治理提供参考。

关键词: 三峡库区; 滑坡; 易发性评价; 机器学习

中图分类号: P642.22

文献标志码: A

文章编号: 1003-8035(2025)03-0140-11

Application of different machine learning models in landslide susceptibility assessment in Badong County, Hubei Province

CHEN Guozhen, LIU Junqi, CUI Jinghui, YU Shupeng, ZHANG Xiaohu, ZHU Ninghui, SONG Yuxin

(Badong National Observation and Research Station of Geohazards, China University of Geosciences,
Wuhan, Hubei 430074, China)

Abstract: China is one of the countries most frequently affected by landslide disasters in the world, making landslide susceptibility assessment crucial for effective disaster prevention and mitigation. Due to variations in the adaptability of different machine learning models in different regions, in order to better carry out landslide disaster prevention and control work in Badong County, Hubei Province, ten influencing factors including slope gradient, slope direction, curvature, degree of undulation, stratigraphy, overburden, NDVI, road density, water system density, and slope structure were selected. 4 different models, including logistic regression, support vector machine, multilayer perceptron, and random forest, were used for landslide susceptibility evaluation. Three evaluation methods were used to assess the accuracy of the model: receiver operating

收稿日期: 2023-10-10; 修订日期: 2024-01-27

投稿网址: <https://www.zgdzhyfzxb.com/>

基金项目: 水利部项目(ZZ2018059); 自然资源部项目(HBHDZFCG2021025)

第一作者: 陈国圳(2001—), 男, 海南海口人, 地学信息工程专业, 硕士研究生, 主要从事地质灾害风险评价研究。

E-mail: 1143470922@qq.com

通讯作者: 刘军旗(1971—), 男, 陕西兴平人, 地学信息工程专业, 副研究员, 主要从事地质灾害大数据技术研究工作。

E-mail: liujqg@qq.com

characteristic curves, mean square error, determination coefficient, and the ratio of landslide to study area. The experimental results show that there are differences among the models in different evaluation methods. Overall, the RF model exhibits the highest accuracy and generates more reasonable susceptibility zoning maps. The susceptibility distribution maps generated by the four models are similar, with high and very high susceptibility areas predominantly located in the southern riverside area. Areas near Guandukou Town and Jiaojiawan Village along the southwest coast exhibit relatively high susceptibility. The assessment results can provide reference for landslide control in Badong County.

Keywords: Three Gorges Reservoir area; landslides; susceptibility assessment; machine learning

0 引言

湖北省三峡库区是全国滑坡灾害发生最频繁的地区之一,库内现共有滑坡 4 664 个,其中 674 个有明显变形特征^[1]。巴东县是三峡库区地质灾害最严重的地区之一,也是全国地质灾害研究的重点地区^[2],开展滑坡易发性评价可为滑坡灾害的预测预警防治提供技术支持,是现阶段滑坡治理的迫切任务。滑坡易发性评价中斜坡单元的划分是重要环节之一,目前划分的方法主要有:水文分析法、地表曲率分水岭法、r.slopeunits 方法等^[3-5]。地质灾害治理得益于过去十多年内人工智能及大数据技术的发展^[6],现机器学习在灾害防治方面得到广泛应用并做出很大的贡献,如刘宝生、王俊德、牟家琦等^[7-11]采用各种模型进行灾害易发性评价,常见的评价模型有:证据权法(weight of evidence, WOE)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、逻辑回归(logical regression, LR)、神经网络(neural networks, NN)、多层感知机(multilayer perceptron, MLP)等^[12-15]。各种模型都有缺陷,如机器学习没有解释个别特征对结果的影响,也没考虑它们的相互依赖性,而深度神经网络缺乏可解释性^[16]。所有的模型都有其特定的性质和缺点,每个模型的性能根据输入数据、模型的结构和准确性而变化^[17]。因此,单一的模型不适用于所有区域,不同地区的同种预测模型产生的结果存在差异,在进行灾害易发性研究时,需根据研究地区的实际条件进行模型的适用性选择。为更好地开展巴东县滑坡防治工作,选取多种机器学习模型进行滑坡易发性评价并对比各个模型精度,找出更适应巴东县地质情况的模型,从而提高滑坡易发性评价的针对性,为巴东县滑坡防治工作提供参考。

针对上述问题,本文通过支持向量机(SVM)、逻辑回归(LR)、多层感知机(MLP)、随机森林(RF)4 种机器学习模型对巴东县进行滑坡易发性评价,获取易发值后将研究区域划分为极高易发区、高易发区、中易发区、

低易发区 4 个区域。并通过受试者工作特征曲线(receiver operating characteristic, ROC)、数理统计等方法对模型精度进行评估。

1 研究区概况

本文研究区位于湖北省巴东县,东至东瀼口镇,西至神农小区,北至谭家屋场,南邻长江,总面积约 10.12 km²。

巴东城区四水环绕、山城镶嵌、地表崎岖、山峦起伏、峡谷幽深、沟壑纵横,是典型的喀斯特地貌^[18]。三峡库区四季分明,气候湿润,年均降雨量近 1 500 mm,水系交错纵横,植被覆盖率高达 55% 以上^[19]。复杂的环境造成巴东县有许多制约发展和危害安全的潜在地质问题,如滑坡、泥石流等^[20-23]。

研究区地势西南高东北低,滑坡密集发育,现共有 25 个滑坡,主要分布于沿江地区。研究区滑坡的高程分布为 142.05 ~ 657.53 m,坡度分布为 0 ~ 63.27°。滑床地层岩性主要为泥质粉砂岩,滑体主要为碎石土。

2 研究方法

2.1 指标及评价因子

网格单元难以反应出地貌和滑坡灾害的关系特征,而斜坡单元与塑造地貌的条件和过程相关,是滑坡发育的基本单元^[24]。本文采用斜坡单元作为评价单元,采用水文分析法进行斜坡单元的划分。

由于滑坡区域常横跨多个斜坡单元,为精确描述滑坡面积,采用 10 m×10 m 的栅格单元,采用栅格数占比代替面积占比。

影响某地区滑坡发生的环境因素十分复杂。因此,很难确定哪些环境因素是最重要和必要的^[25]。刘丽娜等^[26]指出坡度、坡向、曲率是滑坡形成主控因子中的地形因子。李松林^[27]等指出斜坡结构能表现出地形因子的关系,是滑坡发育的关键影响因子之一。石菊松等^[28]指出地表水的冲蚀作用对滑坡有突出影响,水系的密集

程度是滑坡发生的影响因子之一。Wang 等^[29]指出断层边的滑坡运动会相互造成影响,断层距离是常见的滑坡影响因子。但在本研究区内不存在活动断层,断层距离对滑坡的影响几乎可以忽略,因此不作为模型训练的参数。郭慧娟等^[30]指出人类工程活动中修建道路会对两侧的岩土造成影响,诱导滑坡的发生。章昱等^[31]指出植被发育程度直接影响斜坡稳定性。高程体现斜坡的势能和内部应力^[32],而起伏度为高程变化程度,体现了一定范围内势能差,因而陈刚等^[33]指出起伏度是滑坡发生

的指标因子。张玺国等^[34]研究发现地层岩性在影响地质灾害中重要性达 1%~10%。皱浩等^[35]指出覆盖层孕育着斜坡浅表层破坏,是使斜坡从稳定态向不稳定态变化的主要原因,且覆盖层的空间变化圈定滑坡发育的边界。基于前人的研究,结合研究区的地质状况,本文选取坡度、坡向、曲率、起伏度、地层岩性、覆盖层、归一化植被指数、道路密度、水系密度、斜坡结构 10 个可量化提取的影响因子(图 1)。

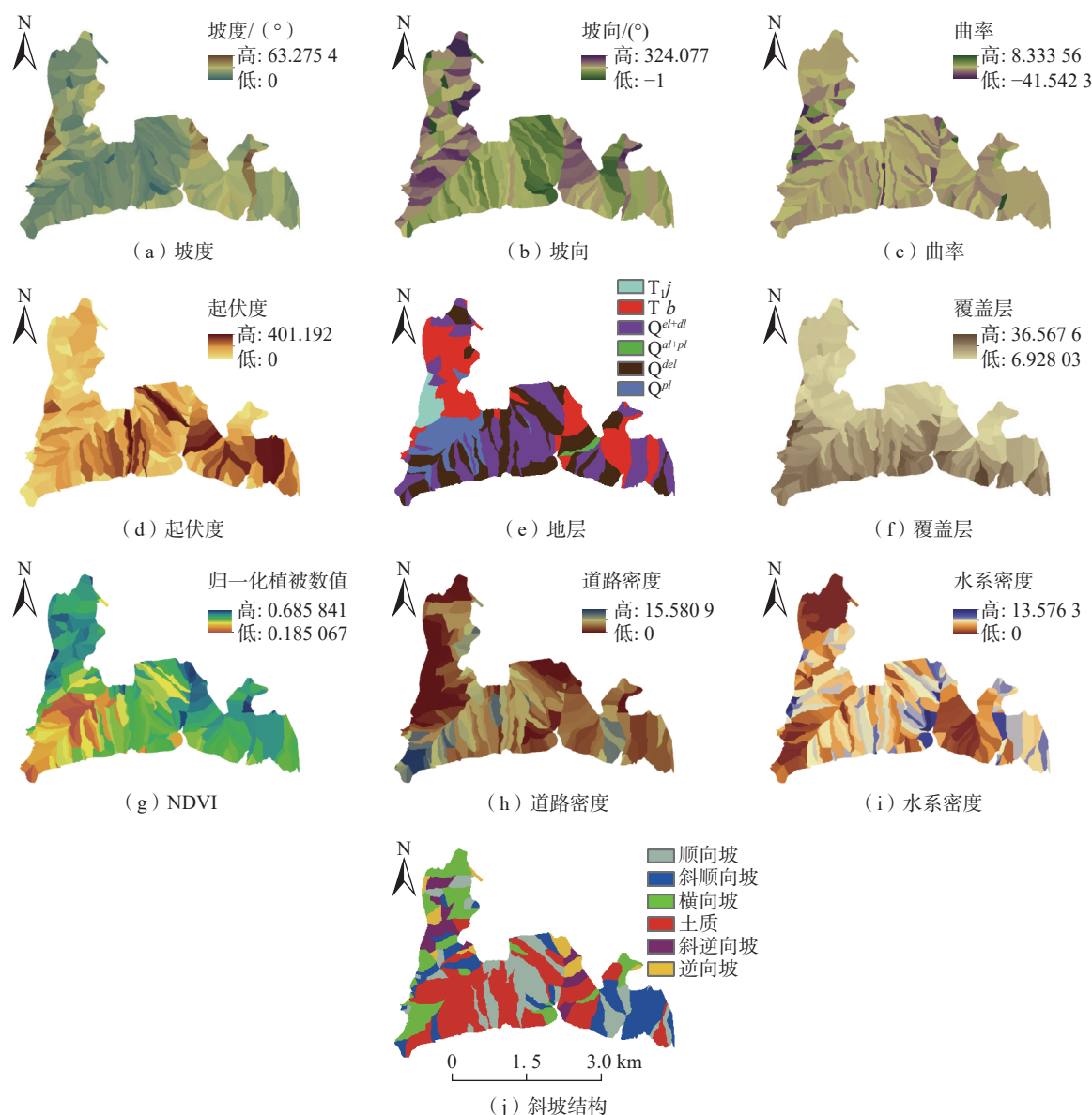


图 1 斜坡易发性评级因子

Fig. 1 Landslide susceptibility rating factor

2.2 机器学习模型

机器学习是滑坡易发性评价常用的模型,本文采用

了 SVM、LR、MLP 和 RF 4 种模型对滑坡易发性进行评价。

(1) SVM

SVM 在解决小样本、非线性和高维模式识别问题中有特有的优势^[36]。对于样本 (x_i, y_i) , $i=1, 2, 3, \dots, n$, 寻找一个超平面:

$$\omega x + b = 0 \quad (1)$$

式中: ω ——决定超平面方向的法向量;

b ——平面和原点之间的位移量。

使得支持向量到超平面(1)的距离最大化, 公式为:

$$d_{\max} = \frac{|\omega x + b|}{\|\omega\|} \quad (2)$$

引入拉格朗日函数, 将问题向高纬度投影, 并使 ω 和 b 偏导数为 0, 可将问题转化为:

$$\min_{\omega, b} L(\omega, b, \alpha) = \sum_{j=1}^n \alpha_j - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (3)$$

式中: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ 可通过 SMO 等算法来获取。

(2) LR

LR 是一种多变量分析模型, 可以根据一组预测变量的值来预测特征或结果的存在与否^[37]。传统的多元线性回归方程:

$$y = \beta_0 + \sum_{i=1}^n x_i \beta_i \quad (4)$$

LR 使用的联系函数为对数几率函数:

$$z = \ln \frac{y}{1-y} \quad (5)$$

将式(4)中的 y 换成 z 带入式(5)中, 可得 LR 的数学原型:

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (6)$$

对式(6)进行变形后, 可得 LR 常用数学公式:

$$y = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (7)$$

式中: β ——回归系数, $\beta = (\beta_0, \beta_1, \dots, \beta_n)$ 可通过极大似然估计法求解。

(3) MLP

MLP 是最常见的全连接人工神经网络(ANN)^[38], MLP 由输入层、隐藏层、输出层 3 层构成^[39]。输入层节点分配权重构建加权线性方程作为下一层的输入, 隐藏层节点值公式为:

$$y = f(\omega x + b) \quad (8)$$

式中: x 、 y ——向量, 隐藏层的数量便是其嵌套数量。

经过向前阶段后, 本文通过 Sigmoid 激

活函数:

$$\text{Sigmoid}(x) = \frac{e^x}{1 + e^x} \quad (9)$$

及其偏导数进行反向传播, 不断更新网络的权值(ω), 使网络缓慢收敛。

(4) RF

RF 是一种集成学习模型(ensemble learning), 其算法是一种 Bagging 算法, 它由多棵决策树组成, 并且由大多数决策树分类结果来决定样本类型^[40]。在回归分析中, 则是取各个决策树的回归结果的平均值作为结果。

2.3 研究思路

本文研究思路(图 2)分为数据获取、模型训练、易发值处理 3 部分。

数据处理部分, 本文所用到的数据均来自资助项目, 从收集到的数据中, 选取 2.1 节中的 10 个指标用于训练。

模型训练部分, 通过选取的特征参数, 生成样本集, 通过测试取得较优的模型参数。其中, SVM 通过 AUC 面积确定参数(C)的较优值; RF 通过均方误差确定较优的决策树数量; MLP 通过多次测试选取较优的层数和节点数。选定较优参数进行模型训练, 以斜坡单元为评价单元, 进行研究区的滑坡易发性评价, 并通过 ROC 曲线和数理统计指标用于评价模型精度。

易发值处理部分, 获取模型预测的易发值后, 剔除异常的易发值, 将易发值归一化后得到处理后的易发值。将处理后的易发值通过自然间断点重分类为 4 类并生成易发分区图, 根据易发性分区图和原有滑坡数据用于数据分析评价模型精度。最后, 将 ROC 曲线、数理统计、数据分析 3 个部分得出的评价结果用于进行综合评价, 得出四个模型最终精度评价结果。

2.4 数据处理

数据处理包含剔除异常值和数据归一化 2 个部分。本文采用的归一化方法为最大-最小标准化, 其公式为:

$$x_{i,\text{new}} = \frac{x_i - x_{\max}}{x_{\max} - x_{\min}} \quad (10)$$

式中: $x_{i,\text{new}}$ ——第 i 个数据归一化后的值;

$x_{\max}$ —— x 数据集中的最大值;

$x_{\min}$ —— x 数据集中的最小值。

在实际应用中, 常出现极少数预测值偏离数据中心, 呈现极大或极小的现象, 造成式(10)中的分母过大, 归一化后大部分预测值极小的情况。本文通过异常斜坡附近的斜坡单元易发性值对其预测, 假设该斜坡单元的序号为 i , 其预测值为:

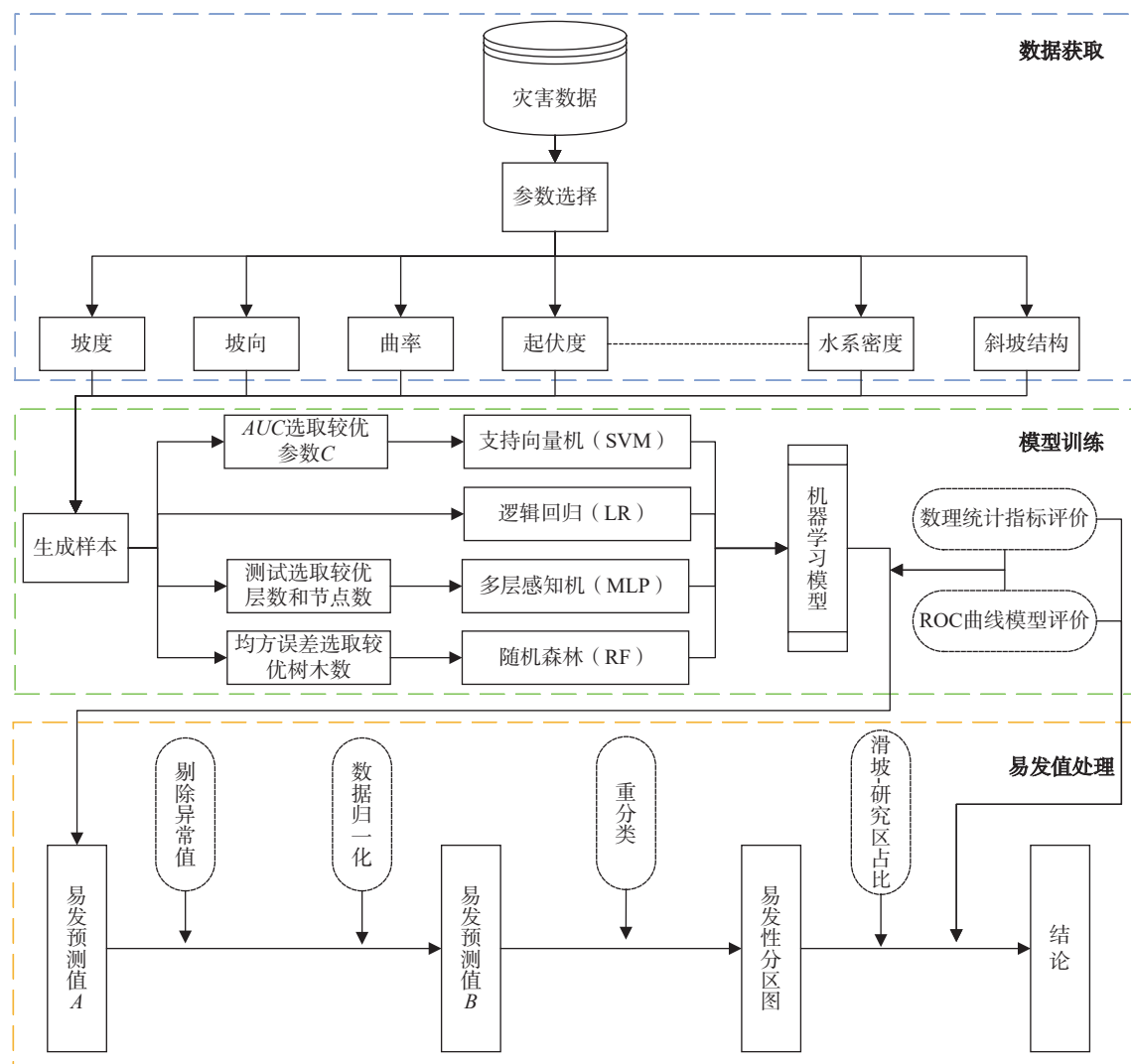


图 2 研究流程图

Fig. 2 Research process flowchart

$$|Value_i| = \begin{cases} \frac{\sum_{n=i+1}^{i+k} |Value_n|}{k}, & i < k \\ \frac{\sum_{n=i-k, n \neq i}^{i+k} |Value_n|}{2k}, & k \leq i \leq size_{\max} - k \\ \frac{\sum_{n=i-k}^{i-1} |Value_n|}{k}, & i > size_{\max} - k \end{cases} \quad (11)$$

式中: $Value$ ——斜坡单元的易发性值;

$size_{\max}$ ——斜坡单元的最大序号;

i 、 n ——斜坡单元序号;

k ——选取邻近斜坡单元的个数。

式(11)表示,当斜坡周围的斜坡单元数量少于 $2k$ 时,选取前或后 k 个斜坡单元的绝对值均值作为预测

值绝对值;若多于 $2k$ 个,则取前 k 个和后 k 个斜坡单元绝对值均值作为预测值绝对值。获得绝对值后,原易发性值的正负作为更新易发值的正负,公式为:

$$Value_i = \begin{cases} |Value_i|, & Value_{\text{old}} \geq 0 \\ -|Value_i|, & Value_{\text{old}} < 0 \end{cases} \quad (12)$$

式中: $Value_i$ ——异常斜坡单元更新后的易发值;

$Value_{\text{old}}$ ——异常斜坡单元的原易发值。

3 研究结果与讨论

本文于灾害区和非灾害区内创建随机点,选取相同数量的灾害点和非灾害点作为样本,样本数量共计 800 个。将样本数量划分为 70% 训练集,30% 测试集两部分。预测区域划分为 245 个斜坡单元,通过 ArcGIS 将斜坡单元与评价因子相结合并转化为预测点集。

本文通过 ROC 曲线、数理统计指标、滑坡-研究区

占比用于模型评估。其中, 滑坡-研究区占比的评价依据是: 通过已知的滑坡区域(灾害区)获取灾害点数据, 由于所选区域本身就是滑坡区域, 即高危险区域, 因此训练出来的模型如果表现为高易发值, 就与实际情况比较符合, 对应模型的模拟效果也就比较好。采用滑坡内部分区占比和滑坡占据研究区不同分区的占比两种占比用于评价。其中, 滑坡内部分区应大部分比例为极高易发区或高易发区, 低易发区比例越少越符合现实情况; 滑坡占据研究区各分区占比中, 滑坡区域在极高易发区或高易发区的占比越高, 低易发区的占比越少视作更加符合现实情况。

3.1 基于 SVM 模型的易发性评价结果

SVM 对参数调节和核函数的选择敏感^[41], 为了使模型更加精确, 在参数 $C \in [1, 500]$ 下绘制其 ROC 曲线, 计算 AUC 面积, AUC 面积代表模型精确度, 面积越接近 1, 模型精度越高^[42]。绘制 AUC 面积与 C 值关系曲线(图 3)。

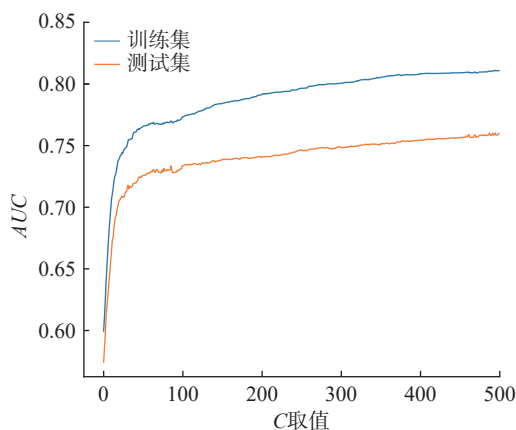


图 3 AUC 面积 - C 值关系图

Fig. 3 Relationship diagram of AUC area - C value

由图 3 可见, 随着 C 值的增大, 模型在训练集和测试集的精确度越来越高, 但测试集精确度在 $C=200$ 后增大不再显著, 为防止模型过拟合与减少复杂度, 选择参数 C 的值为 200 进行训练。将训练后的模型用于对斜坡单元进行预测, 预测的易发性值进行剔除异常值、归一化后通过 ArcGIS 进行重分类得到易发性评价分区图(图 4)。

计算滑坡内部分区的栅格数占比用于模型精确度分析。滑坡区域中极高易发区(34.73%)和高易发区(24.63%)的占比为 59.36%, 超过一半的滑坡区域为高易发值区域, 中易发区(27.98%)和低易发区(12.66%)2 区占比达到 40.64%, 可见预测结果与真实值稍有偏差。

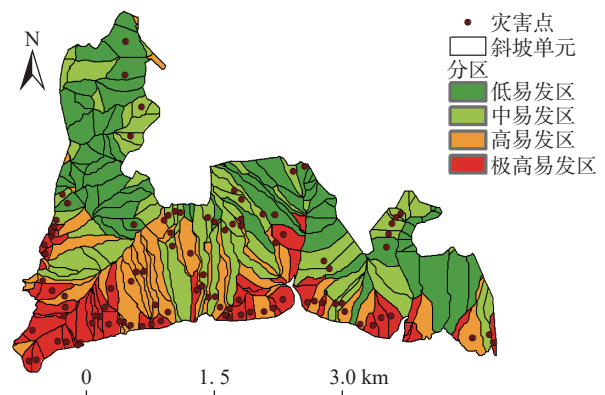


图 4 基于 SVM 滑坡易发性评价分区图

Fig. 4 Landslide susceptibility map based on SVM

3.2 基于 LR 模型的易发性评价结果

LR 实际上是检查独立因素对二元结果的贡献的系统且强大的模型^[43]。通过训练模型获得独立因素对分类结果的贡献值, 即式(7)中的 $(\beta_1, \beta_2, \dots, \beta_n)$, 式中的自变量 $(x_1, x_2, \dots, x_n)$ 为斜坡易发性评级因子(图 1)。根据式(7)通过 ArcGIS 的“栅格计算器”功能, 可计算得到斜坡易发性分区图(图 5)。

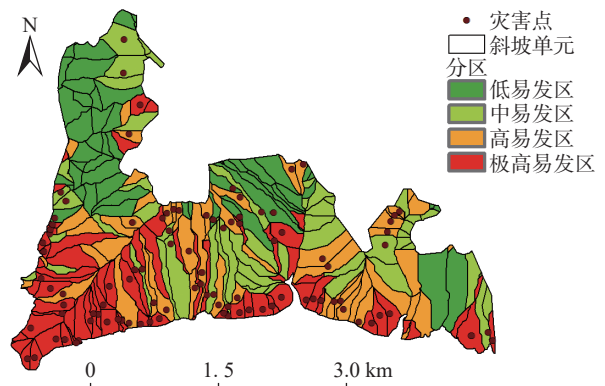


图 5 基于 LR 滑坡易发性评价分区图

Fig. 5 Landslide susceptibility map based on LR

计算滑坡内部分区的栅格数占比用于模型精确度分析。滑坡区域中极高易发区占比 46.13%, 高易发区占比 32.60%, 2 区占比共计 78.73%, 中易发区 15.03%, 低易发区 6.24%, 2 区占比共计 21.27%。滑坡区域中大部分为高易发值区域, 小部分为中低易发区, 可见模型精度较高。

3.3 基于 MLP 模型易发性评价结果

MLP 隐藏层的层数与节点数难以找到最佳的结果。在通过多次测试后, 选择层数为 5, 节点数分别为 160, 125, 90, 75, 40 的结果较稳定良好。通过模型训练后进行易发值预测, 最后对易发性图进行重分类可得易发性评价分区图(图 6)。

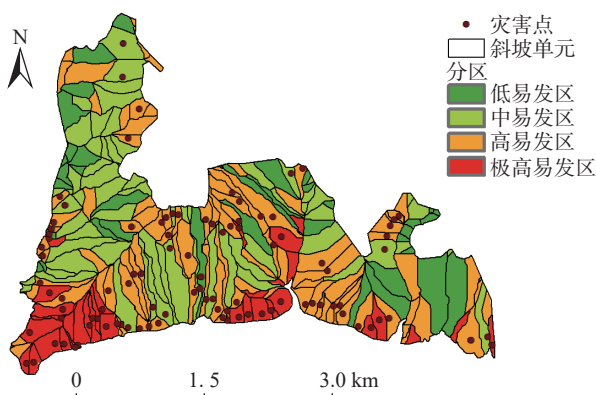


图 6 基于 MLP 滑坡易发性评价分区图

Fig. 6 Landslide susceptibility map based on MLP

计算滑坡内部分区的栅格数占比用于模型精确度分析。滑坡区域中极高易发区比例为 26.58%, 高易发区占 52.09%, 2 区共计 78.67%, 低易发区仅占 3.31%, 中低易发 2 区共占 21.33%。滑坡区域中的中高易发区较高, 低易发值只占小部分, 由此可见模型精度较高。

3.4 基于 RF 模型的易发性评价结果

RF 的调参是一个关键步骤, 既要建立优良性能的模型, 又要防止模型过拟合^[44]。其中, 最为重要的参数是森林中树木的数量, 树木的数量越多模型的效果往往越好, 但决策树数量过多会出现过拟合问题, 因此需先寻找合适的决策树数量。模型精确度可通过训练集和测试集的均方误差体现。现绘制训练集和测试集的均方误差和树木数量关系曲线图(图 7)。

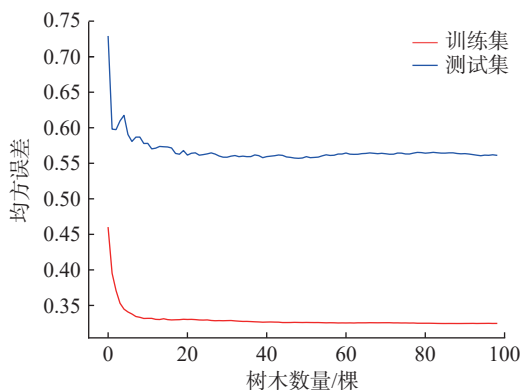


图 7 均方误差-树木数量关系图

Fig. 7 Relationship between mean square error and number of trees

由图 7 可见, 在 1~100 的范围内, 随着树木数量增加, 模型精度不断提高, 且测试集的均方误差趋近平稳, 证明模型没有过拟合。在树木数量为 40 之后, 精度变化不显著, 继续增加树木数量会增大算法复杂度, 因此可取树木数量为 40 的模型用于预测。将预测结果通过图 1 中的易发值处理流程后得到易发性分区图(图 8)。

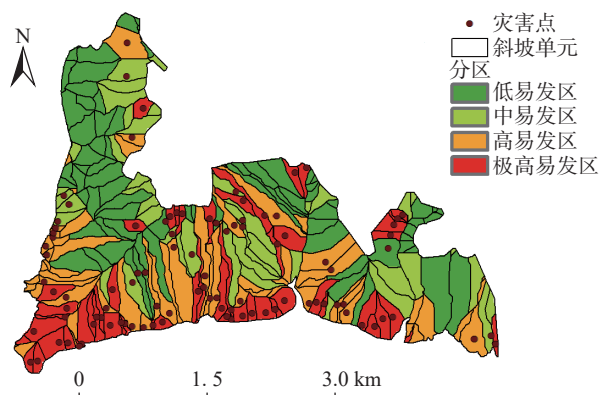


图 8 基于 RF 滑坡易发性评价分区图

Fig. 8 Landslide susceptibility map based on RF

计算滑坡内部分区的栅格数占比用于模型精确度分析。滑坡区域中极高易发区占 52.77%, 高易发区占 34.34%。2 区共计 87.11%, 中易发区 9.7%, 且低易发区占比低至 3.19%, 滑坡区域中高易发值区域占比接近 90%, 中低易发值区域占比仅为 12.89%, 可见模型精度非常高。

3.5 结果讨论

本文通过 4 种模型对巴东县滑坡易发性进行了评价, 为分析易发性分区评价结果, 统计 4 种模型分区占比(表 1), MLP 的评价结果在高易发区和中易发区的覆盖率较高, 均在 30% 以上, 而极高易发区显著低于其他模型; SVM 的评价结果较为保守, 中低易发区比例相对较高, 覆盖率共达 64.08%; LR 的各分区分布比例较为均匀, 均在 20%~30%; RF 的低易发区比例较高, 其他 3 个分区的比例相近。

表 1 各模型评价结果分区占比表

Table 1 Proportion of landslide zone assessment results for each model

	极高/%	高/%	中/%	低/%
LR	26.36	23.98	24.29	25.37
SVM	16.53	19.39	29.34	34.74
MLP	12.24	36.61	33.04	18.11
RF	20.16	25.58	20.75	33.51

观察各模型的易发性评价结果图(图 4—6、8)的各分区分布情况, 可见 4 种模型的评价结果相近, 极高易发区和高易发区都大部分分布于西南的长江沿岸地区, 部分区域沿着南岸分布, 低易发区主要分布于西北区域。极高易发区和高易发区都主要分布于官渡口镇、焦家湾村等沿岸地区附近。相似的地区分布结果表明各模型对滑坡易发性评价的一致性和较高可信度。

通过各模型的 ROC 曲线(图 9)对 4 种模型进行对比,

ROC 下的 AUC 可表示模型精确度。LR 模型的 $AUC=0.74$, SVM 模型的 $AUC=0.76$, MLP 模型的 $AUC=0.73$, RF 模型的 $AUC=0.91$ 。在 AUC 值方面, 所有模型都取得良好的模型性能 ($AUC>0.70$), 可见 4 种模型都取得较好的效果。其中, RF 对滑坡易发性的评价效果最佳, 达到了 0.90 以上, 其他 3 种模型的评价效果相似, 均分布在 0.70 ~ 0.80。

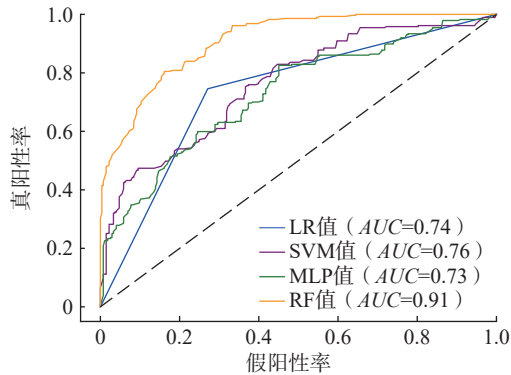


图9 模型 ROC 曲线图

Fig. 9 Comparison of ROC curves for different models

考虑到不同模型的不确定性影响, 进一步通过数理统计方法来判断 4 种回归模型中预测值和真实值的差异, 常用指标为均方误差 (mean square error, MSE)、决定系数 (coefficient of determination, CD)、平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean squared error, $RMSE$), 本文将通过上述 4 项指标评价模型在数据上的拟合程度。计算各模型的 4 项指标 (图 10), 根据上述指标介绍的顺序排序, LR 的各项指标为 1.06、-0.06、0.53、1.03, SVM 的各项指标为 0.86、0.14、0.73、0.93, MLP 的各项指标为 2.60、-1.60、0.98、1.61, RF 的各项指标为 0.51、0.49、0.59、0.72。其中, CD 值越接近 1 表明预测值和真实值的差异越小, MSE 、 MAE 、 $RMSE$ 值越接近 0 表明预测值和真实值差异越小。由此可见, MLP 的 CD 值和 MSE 值与其他模型的值偏差较多, 预测结果和真实值差异性较大, 而 RF 的预测结果优于其他 3 个模型。

在特殊情况下, 模型预测的易发值过度偏大, 造成极高易发区和高易发区范围比例过度偏高, 滑坡内部分区占比的评价依据发生误判。为使滑坡-研究区占比的评价方式更加精确, 统计滑坡区域占据研究区不同分区的占比 (图 11), 即滑坡占据各易发区的比例。从滑坡占据研究区各分区的比例观察, 训练的模型在灾害区应表现出高易发值, 理想情况下灾害区域应分布在极高易发区, 因此滑坡地区占据极高易发区的比例越高越符

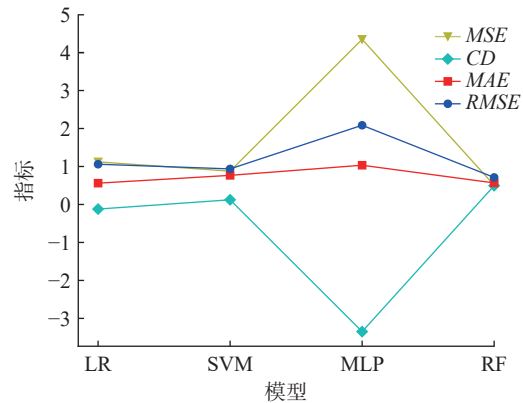


图 10 数理统计指标图

Fig. 10 Statistical metrics chart

合现实情况, 高易发区次之; 滑坡区域若是出现在低易发值区域则不合理, 因此占据低易发区的比例越高越不符合现实情况。总体而言, 滑坡区域极大地占据 RF 预测结果中的极高易发区, 占据比例接近 70%, 而 LR 的滑坡区域在高易发值区占比稍低, 但 SVM、LR 和 MLP 的结果精度总体相近。由此可见, RF 的预测结果最佳, SVM、LR 和 MLP 的结果相近, 在 3 种评价方法上都略低于 RF 模型。

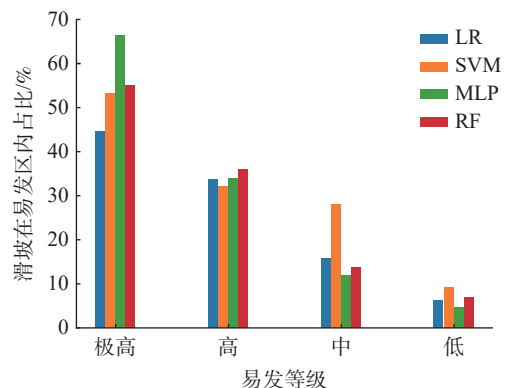


图 11 滑坡占据各易发区占比

Fig. 11 Landslide - susceptibility area proportion

4 结论

(1) 4 种模型的评价结果显示高易发值区主要分布在南边沿江地区, 低易发值区主要分布于西北区域, 西南沿岸的官渡口镇、焦家湾村等附近地区展现出较高易发性, 相似的评价结果表明评价的一致性和较高的可信度。

(2) 试验表明, 在 AUC 方面, RF 有显著优势 ($AUC>0.9$), 其余 3 个模型相近 ($0.8>AUC>0.7$), 可见 RF 有着较优的模型精度; 在预测值和实际值的近似度方面,

MLP 表现出较大偏差;在高易发值区域和实际滑坡区域的相似度方面,SVM 的相似度最低,滑坡区域的中低易发区比例达到 40.64%;LR 的极高易发区中滑坡的占比为 44.65%,略低于其他模型。不同模型在不同评价方式中存在差异,但总体可见,RF 模型精度最高且绘制出的易发性分区图更合理。

(3) 通过不同机器学习模型在巴东城区的滑坡易发性评价应用研究,可以为三峡库区类似的重点城镇滑坡易发性评价工作提供参考范例,为三峡库区重点城镇的规划及可持续发展提供技术支持。

参考文献(References):

- [1] 吴宏阳,周超,梁鑫,等.基于 XGBoost 模型的三峡库区燕山乡滑坡易发性评价与区划[J].中国地质灾害与防治学报,2023,34(5):141-152. [WU Hongyang, ZHOU Chao, LIANG Xin, et al. Assessment of landslide susceptibility mapping based on XGBoost model: A case study of Yanshan Township [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(5): 141-152. (in Chinese with English abstract)]
- [2] 陈国金.三峡库区巴东城区岸坡稳定性问题调查研究与风险控制[J].资源环境与工程,2023,37(3):292-300. [CHEN Guojin. Stability investigation and risk control of bank slope in Badong urban area, Three Gorges Reservoir area [J]. Resources Environment & Engineering, 2023, 37(3): 292-300. (in Chinese with English abstract)]
- [3] 张曦,陈丽霞,徐勇,等.两种斜坡单元划分方法对滑坡灾害易发性评价的对比研究[J].安全与环境工程,2018,25(1):12-17. [ZHANG Xi, CHEN Lixia, XU Yong, et al. Comparison of two methods for slope unit division in landslide susceptibility evaluation [J]. Safety and Environmental Engineering, 2018, 25(1): 12-17. (in Chinese with English abstract)]
- [4] 曾斌,吕权儒,寇磊,等.基于 Logistic 回归和随机森林的清江流域长阳库岸段堆积层滑坡易发性评价[J].中国地质灾害与防治学报,2023,34(4):105-113. [ZENG Bin, LYU Quanru, KOU Lei, et al. Susceptibility assessment of colluvium landslides along the Changyang section of Qingjiang River using logistic regression and random forest methods [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(4): 105-113. (in Chinese with English abstract)]
- [5] 王凯,张少杰,韦方强.斜坡单元提取方法研究进展和展望[J].长江科学院院报,2020,37(6):85-93. [WANG Kai, ZHANG Shaojie, WEI Fangqiang. Slope unit extraction methods: Advances and prospects [J]. Journal of Yangtze River Scientific Research Institute, 2020, 37(6): 85-93. (in Chinese with English abstract)]
- [6] 刘军旗,刘强,刘千慧,等.大数据时代地质灾害数据管理及应用模式探讨[J].地质科技通报,2021,40(6):276-282. [LIU Junqi, LIU Qiang, LIU Qianhui, et al. Discussion of geological hazard data management and application model in big data era [J]. Bulletin of Geological Science and Technology, 2021, 40(6): 276-282. (in Chinese with English abstract)]
- [7] 刘宝生,陈刚,程刚建.江苏南京地质灾害风险评价[J].中国地质灾害与防治学报,2023,34(4):97-104. [LIU Baosheng, CHEN Gang, CHENG Gangjian. Risk assessment of geological disasters in Nanjing, Jiangsu Province [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(4): 97-104. (in Chinese with English abstract)]
- [8] 王俊德,杜晓阳,黄天浩,等.河南省嵩县地质灾害风险评价[J].中国地质灾害与防治学报,2023,34(4):86-96. [WANG Junde, DU Xiaoyang, HUANG Tianhao, et al. Risk assessment of geological hazards in Song County, Henan Province [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(4): 86-96. (in Chinese with English abstract)]
- [9] 牟家琦,庄建琦,王世宝,等.基于深度神经网络模型的雅安市滑坡易发性评价[J].中国地质灾害与防治学报,2023,34(3):157-168. [MU Jiaqi, ZHUANG Jianqi, WANG Shibao, et al. Evaluation of landslide susceptibility in Ya'an City based on depth neural network model [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(3): 157-168. (in Chinese with English abstract)]
- [10] 支泽民,刘峰贵,周强,等.基于流域单元的地质灾害易发性评价——以西藏昌都市为例[J].中国地质灾害与防治学报,2023,34(1):139-150. [ZHI Zemin, LIU Fenggui, ZHOU Qiang, et al. Evaluation of geological hazards susceptibility based on watershed units: A case study of the Changdu City, Tibet [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(1): 139-150. (in Chinese with English abstract)]
- [11] 郭飞,王秀娟,陈玺,等.基于不同模型的赣南地区小型削方滑坡易发性评价对比分析[J].中国地质灾害与防治学报,2022,33(6):125-133. [GUO Fei, WANG Xiujuan, CHEN Xi, et al. Comparative analyses on susceptibility of cutting slope landslides in southern Jiangxi using different models [J]. The Chinese Journal of Geological Hazard and Control, 2022, 33(6): 125-133. (in Chinese with English abstract)]
- [12] 黄发明,殷坤龙,蒋水华,等.基于聚类分析和支持向量机的滑坡易发性评价[J].岩石力学与工程学报,2018,37(1):156-167. [HUANG Faming, YIN Kunlong, JIANG Shuihua, et al. Landslide susceptibility assessment based on

- clustering analysis and support vector machine [J]. Chinese Journal of Rock Mechanics and Engineering, 2018, 37(1): 156 – 167. (in Chinese with English abstract)]
- [13] WANG Jiaqi, SUN Pengfei, CHEN Leilei, et al. Recent advances of deep learning in geological hazard forecasting [J]. Computer Modeling in Engineering & Sciences, 2023, 137(2): 1381 – 1418.
- [14] 夏辉,殷坤龙,梁鑫,等.基于SVM-ANN模型的滑坡易发性评价——以三峡库区巫山县为例 [J].中国地质灾害与防治学报,2018,29(5):13 – 19. [XIA Hui, YIN Kunlong, LIANG Xin, et al. Landslide susceptibility assessment based on SVM-ANN Models: A case study for Wushan County in the Three Gorges Reservoir [J]. The Chinese Journal of Geological Hazard and Control, 2018, 29(5): 13 – 19. (in Chinese with English abstract)]
- [15] 何清,李宁,罗文娟,等.大数据下的机器学习算法综述 [J].模式识别与人工智能,2014,27(4):327 – 336. [HE Qing, LI Ning, LUO Wenjuan, et al. A survey of machine learning algorithms for big data [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327 – 336. (in Chinese with English abstract)]
- [16] YOUSSEF K, SHAO K, MOON S, et al. Landslide susceptibility modeling by interpretable neural network [J]. Communications Earth and Environment, 2023, 4(1): 162.
- [17] NACHAPPA T, GHORBANZADEH O, GHOLAMNIA K, et al. Multi-hazard exposure mapping using machine learning for the state of Salzburg, Austria [J]. Remote Sensing, 2020, 12(17): 2757.
- [18] 马晨曦.后三峡时期库区城市人居环境建设评价研究——以巴东、秭归为例 [D].重庆:重庆大学,2019. [MA Chenxi. Evaluation of urban human settlement environment construction in the reservoir area in the post-Three Gorges period [D]. Chongqing: Chongqing University, 2019. (in Chinese with English abstract)]
- [19] 左可顺.高切坡数据集成与高效管理研究——以三峡湖北库区为例 [D].武汉:中国地质大学(武汉),2023. [ZUO Keshun. Research on data integration and efficient management of high cutting slope: Taking Three Gorges Hubei Reservoir area as an example [D]. Wuhan: China University of Geosciences (Wuhan), 2023. (in Chinese with English abstract)]
- [20] 石菊松,张永双,董诚,等.基于GIS技术的巴东新城区滑坡灾害危险性区划 [J].地球学报,2005,26(3):275 – 282. [SHI Jusong, ZHANG Yongshuang, DONG Cheng, et al. GIS-based landslide hazard zonation of the new Badong County site [J]. Acta Geoscientia Sinica, 2005, 26(3): 275 – 282. (in Chinese with English abstract)]
- [21] 王瑛,林齐根,史培军.中国地质灾害伤亡事件的空间格局及影响因素 [J].地理学报,2017,72(5):906 – 917. [WANG Ying, LIN Qigen, SHI Peijun. Spatial pattern and influencing factors of casualty events caused by landslides [J]. Acta Geographica Sinica, 2017, 72(5): 906 – 917. (in Chinese with English abstract)]
- [22] 吴润泽,程温鸣,刘军旗,等.三峡库区地质灾害防治信息系统及预警指挥系统数据管理模式探讨 [J].中国地质灾害与防治学报,2018,29(5):102 – 107. [WU Runze, CHENG Wenming, LIU Junqi, et al. Discussion on the data management mode of geologic disaster prevention and control information system and early warning command system in the Three Gorges Reservoir area [J]. The Chinese Journal of Geological Hazard and Control, 2018, 29(5): 102 – 107. (in Chinese with English abstract)]
- [23] LIU Junqi, HUANG Xuebin, WU Chonglong, et al. From the area to the point-study on the key technology of 3D geological hazard modeling in Three Gorges Reservoir area [J]. Journal of Earth Science, 2012, 23(2): 199 – 206.
- [24] 刘超凡,付萧,朱庆,等.顾及地貌形态特征的精细斜坡单元高效分区划分 [J].测绘科学,2023,48(4):211 – 220. [LIU Yuefan, FU Xiao, ZHU Qing, et al. An efficient and fine slope units division method with consideration of the regional geomorphological characteristics [J]. Science of Surveying and Mapping, 2023, 48(4): 211 – 220. (in Chinese with English abstract)]
- [25] HUANG Faming, CAO Zhongshan, JIANG Shuihua, et al. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model [J]. Landslides, 2020, 17(12): 2919 – 2930.
- [26] 刘丽娜,许冲,徐锡伟,等.GIS支持下基于AHP方法的2013年芦山地震区滑坡危险性评价 [J].灾害学,2014,29(4):183 – 191. [LIU Lina, XU Chong, XU Xiwei, et al. GIS-based landslide hazard evaluation using AHP method in the 2013 Lushan earthquake region [J]. Journal of Catastrophology, 2014, 29(4): 183 – 191. (in Chinese with English abstract)]
- [27] 李松林,许强,汤明高,等.三峡库区滑坡空间发育规律及其关键影响因子 [J].地球科学,2020,45(1):341 – 354. [LI Songlin, XU Qiang, TANG Minggao, et al. Study on spatial distribution and key influencing factors of landslides in Three Gorges Reservoir area [J]. Earth Science, 2020, 45(1): 341 – 354. (in Chinese with English abstract)]
- [28] 石菊松,徐瑞春,石玲,等.基于RS和GIS技术的清江隔河岩库区滑坡易发性评价与制图 [J].地学前缘,2007,14(6):119 – 128. [SHI Jusong, XU Ruichun, SHI Ling, et al. ETM+ imagery and GIS-based landslide susceptibility mapping for the regional area of Geheyan Reservoir on the Qingjiang River, Hubei Province, China [J]. Earth Science Frontiers,

- 2007, 14(6): 119–128. (in Chinese with English abstract)]
- [29] WANG Jing, XIANG Wei, LU Ning. Landsliding triggered by reservoir operation: A general conceptual model with a case study at Three Gorges Reservoir [J]. *Acta Geotechnica*, 2014, 9(5): 771–788.
- [30] 郭惠娟, 唐南奇, 林金宝. 基于 GIS 的仙游县土地利用与滑坡灾害敏感性分析 [J]. 福建农林大学学报(自然科学版), 2010, 39(4): 417–420. [GUO Huijuan, TANG Nanqi, LIN Jinbao. Sensibility analysis of land-use and landslide hazard based on GIS in Xianyou County [J]. *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 2010, 39(4): 417–420. (in Chinese with English abstract)]
- [31] 章昱, 王磊, 伏永朋, 等. 基于斜坡单元与信息量法的丹江口库区典型流域地质灾害易发性评价 [J]. *华南地质*, 2023, 39(3): 512–522. [ZHANG Yu, WANG Lei, FU Yongpeng, et al. Evaluation of geological disaster susceptibility in typical watershed of Danjiangkou Reservoir area based on slope unit and information method [J]. *South China Geology*, 2023, 39(3): 512–522. (in Chinese with English abstract)]
- [32] 刘亮, 杨洋. 基于网格单元和随机森林的滑坡易发性评价 [J]. *石化技术*, 2023, 30(6): 180–182. [LIU Liang, YANG Yang. Landslide susceptibility evaluation based on grid cells and random forest [J]. *Petrochemical Industry Technology*, 2023, 30(6): 180–182. (in Chinese with English abstract)]
- [33] 陈刚, 郝社锋, 蒋波等. 基于机载 LiDAR 技术植被茂密区小型滑坡识别与评价 [J]. *自然资源遥感*, 2023, 36(3): 196–205. [CHEN Gang, HE Shefeng, JIANG Bo, et al. Identification and assessment of small landslides in densely vegetated areas based on airborne LiDAR technology [J]. *Remote Sensing of Natural Resources*, 2023, 36(3): 196–205. (in Chinese with English abstract)]
- [34] 张玺国, 周雄冬, 徐梦珍, 等. 西藏地质灾害易发性及对水能开发适宜度影响 [J]. *地理学报*, 2022, 77(7): 1603–1614. [ZHANG Xiguo, ZHOU Xiongdong, XU Mengzhen, et al. Distribution of hydropower development suitability in Tibet in the face of geological hazard susceptibility [J]. *Acta Geographica Sinica*, 2022, 77(7): 1603–1614. (in Chinese with English abstract)]
- [35] 邹浩, 贾琳, 郑路路, 等. 基于覆盖土层厚度识别的区域斜坡降雨入渗稳定性定量评价 [J]. *地球科学*, 2023, 49(9): 3347–3362. [ZHOU Hao, JIA Lin, ZHENG Lulu, et al. Regional Hillslope Stability Analysis under Rainfall Based on Characterization of Overburden Soil Layer Thickness [J]. *Earth Science*, 2023, 49(9): 3347–3362. (in Chinese with English abstract)]
- [36] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述 [J]. *电子科技大学学报*, 2011, 40(1): 2–10. [DING Shifei, QI Bingjuan, TAN Hongyan. An overview on theory and algorithm of support vector machines [J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(1): 2–10. (in Chinese with English abstract)]
- [37] WANG Sen, LING Sixiang, WU Xiyong, et al. Key predisposing factors and susceptibility assessment of landslides along the Yunnan–Tibet traffic corridor, Tibetan Plateau: Comparison with the LR, RF, NB, and MLP techniques [J]. *Frontiers in Earth Science*, 2023, 10: 1100363.
- [38] YU Haiwei, PEI Wenjie, ZHANG Jingyi, et al. Landslide susceptibility mapping and driving mechanisms in a vulnerable region based on multiple machine learning models [J]. *Remote Sensing*, 2023, 15(7): 1886.
- [39] RAGHU S, SRIRAM N. Optimal configuration of multilayer perceptron neural network classifier for recognition of intracranial epileptic seizures [J]. *Expert Systems with Applications*, 2017, 89(C): 205–221.
- [40] TASER P Y. Application of bagging and boosting approaches using decision tree-based algorithms in diabetes risk prediction [C]. //The 7th International Management Information Systems Conference. Basel Switzerland: MDPI, 2021, 89(C): 205–221.
- [41] 窦杰, 向子林, 许强, 等. 机器学习在滑坡智能防灾减灾中的应用与发展趋势 [J]. *地球科学*, 2023, 48(5): 1657–1674. [DOU Jie, XIANG Zilin, XU Qiang, et al. Application and development trend of machine learning in landslide intelligent disaster prevention and mitigation [J]. *Earth Science*, 2023, 48(5): 1657–1674. (in Chinese with English abstract)]
- [42] 孙滨, 祝传兵, 康晓波, 等. 基于信息量模型的云南东川泥石流易发性评价 [J]. *中国地质灾害与防治学报*, 2022, 33(5): 119–127. [SUN Bin, ZHU Chuanbing, KANG Xiaobo, et al. Susceptibility assessment of debris flows based on information model in Dongchuan, Yunnan Province [J]. *The Chinese Journal of Geological Hazard and Control*, 2022, 33(5): 119–127. (in Chinese with English abstract)]
- [43] NAMAN KAUR, HIMANSHU. Logistic regression: A basic approach [J]. *Information and Communication Technology for Competitive Strategies*, 2022, 623: 481–488.
- [44] 吴润泽, 胡旭东, 梅红波, 等. 基于随机森林的滑坡空间易发性评价——以三峡库区湖北段为例 [J]. *地球科学*, 2021, 46(1): 321–330. [WU Runze, HU Xudong, MEI Hongbo, et al. Spatial susceptibility assessment of landslides based on random forest: A case study from Hubei section in the Three Gorges Reservoir area [J]. *Earth Science*, 2021, 46(1): 321–330. (in Chinese with English abstract)]