

非滑坡样本选择对滑坡易发性评价的影响研究

龚 屿, 刘 晓

Analyzing the influence of non-landslide sample selection on landslide susceptibility: Case studies from Wenchuan, Lixian and Maoxian Counties

GONG Yu and LIU Xiao

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.202401009>

您可能感兴趣的其他文章

Articles you may be interested in

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

基于信息量、加权信息量与逻辑回归耦合模型的云南罗平县崩滑灾害易发性评价对比分析

Comparative analyses of susceptibility assessment for landslide disasters based on information value, weighted information value and logistic regression coupled model in Luoping County, Yunnan Province

杨得虎, 朱杰勇, 刘帅, 马博, 代旭升 中国地质灾害与防治学报. 2023, 34(5): 43–53

基于快速聚类-信息量模型的汶川及周边两县滑坡易发性评价

Landslide susceptibility assessment based on K-means cluster information model in Wenchuan and two neighboring counties, China

周天伦, 曾超, 范晨, 毕鸿基, 龚恩慧, 刘晓 中国地质灾害与防治学报. 2021, 32(5): 137–150

基于不同耦合模型的县域滑坡易发性评价对比分析

Comparison of landslide susceptibility assessment based on multiple hybrid models at county level: A case study for Puge County, Sichuan Province

熊小辉, 汪长林, 白永健, 铁永波, 高延超, 李光辉 中国地质灾害与防治学报. 2022, 33(4): 114–124

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

基于RBF神经网络信息量耦合模型的滑坡易发性评价

Landslide susceptibility assessment by the coupling method of RBF neural network and information value: A case study in Min Xian, Gansu Province

黄立鑫, 郝君明, 李旺平, 周兆叶, 贾佩钱 中国地质灾害与防治学报. 2021, 32(6): 116–126



关注微信公众号，获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.202401009

龚屿, 刘晓. 非滑坡样本选择对滑坡易发性评价的影响研究——以汶川县、理县和茂县为例[J]. 中国地质灾害与防治学报, 2025, 36(3): 129-139.

GONG Yu, LIU Xiao. Analyzing the influence of non-landslide sample selection on landslide susceptibility: Case studies from Wenchuan, Lixian and Maoxian Counties[J]. The Chinese Journal of Geological Hazard and Control, 2025, 36(3): 129-139.

非滑坡样本选择对滑坡易发性评价的影响研究 ——以汶川县、理县和茂县为例

龚 屿, 刘 晓

(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: 研究探索了机器学习在评估滑坡易发性中的应用, 重点关注非滑坡样本的选择问题。以四川省汶川县、理县和茂县为研究区, 选取坡度、坡向、高程、距水系距离、距断层距离、岩性和土地利用 7 个评价因子, 从信息量模型(I)、证据权重模型(WOE)、确定性系数模型(CF)、频率比模型(FR)划分的较低和极低易发区以及缓冲区外(B)和全区(G)随机选取非滑坡样本, 构建基于不同非滑坡样本选取方法的支持向量机模型(SVM)并展开易发性评价。结果显示: I-SVM、WOE-SVM、CF-SVM、FR-SVM 的 ROC 曲线下 AUC 值分别为 0.980 4、0.972 6、0.936 8、0.845 1, 优于 B-SVM 的 0.786 9 和 G-SVM 的 0.738 9, 说明采用数学统计模型所选取的非滑坡样本准确性更高, 信息量模型是选取非滑坡样本的最优方法, 为非滑坡样本的选取提供新思路。

关键词: 滑坡; 易发性评价; 信息量模型; 证据权; 确定性系数; 频率比; 支持向量机

中图分类号: P642.22

文献标志码: A

文章编号: 1003-8035(2025)03-0129-11

Analyzing the influence of non-landslide sample selection on landslide susceptibility: Case studies from Wenchuan, Lixian and Maoxian Counties

GONG Yu, LIU Xiao

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

Abstract: This research explores the integration of machine learning in assessing landslide susceptibility, scrutinizing the selection of non-landslide samples. Taking Wenchuan County, Lixian County, and Maoxian County in Sichuan Province as the study areas, 7 evaluation factors were considered, including slope, aspect, elevation, distance to the water system, distance to the fault, lithology, and land use. Non-landslide samples were randomly selected from the lower and extremely low susceptibility zones divided by the information value model (I), weight of evidence model(WOE), coefficient of determination model (CF),

收稿日期: 2024-01-04; 修订日期: 2024-03-05

投稿网址: <https://www.zgdzhyfzxb.com/>

基金项目: 国家自然科学基金面上项目(42072314; 41572279); 中国博士后科学基金特别资助项目(2014T70758); 中国博士后科学基金面上资助项目(2012M521500); 中交第二公路勘察设计研究院有限公司科技研发项目(KJFZ-2018-049)

第一作者: 龚 屿(1998—), 男, 湖北仙桃人, 地质工程专业, 硕士, 主要从事地质工程与地质灾害方面的研究。

E-mail: 13007133876@163.com

通讯作者: 刘 晓(1977—), 男, 湖北武汉人, 岩土工程专业, 博士, 副研究员, 主要从事工程地质模拟、滑坡演化机制和控制理论方面的研究。E-mail: liuxiao@china.com

and frequency ratio model(FR), as well as form the buffer zones (B) and the entire region (G). These samples were then analyzed using a support vector machine (SVM) model. The results showed that the AUC values for I-SVM, WOE-SVM, CF-SVM, and FR-SVM were 0.980 4, 0.972 6, 0.936 8, and 0.845 1, respectively, which were superior to the AUC values of B-SVM (0.786 9) and G-SVM (0.738 9). This highlight the effectiveness of using mathematical-statistical models for the selection of non-landslide samples, with particular emphasis on the accuracy of the information value model. This study offers a novel approach to selecting non-landslide samples, significantly enhancing predictive accuracy in landslide susceptibility assessments.

Keywords: landslide; susceptibility assessment; information value model; weight of evidence; certainty factor; frequency ratio; support vector machine

0 引言

滑坡易发性评价是预防地质灾害、确保人类生命财产安全的重要一环。近年来,机器学习模型被大量应用于滑坡易发性评价,例如支持向量机模型(support vector machine, SVM)^[1]、决策树模型(decision tree, DT)^[2]、随机森林模型(random forest, RF)^[3]、人工神经网络(artificial neural networks, ANN)^[4]等。相比于传统的数学统计模型,机器学习模型精度更高,但模型性能高度依赖于样本质量,除了滑坡样本(正样本),还需要选取非滑坡样本(负样本)。

对于非滑坡点的选取,目前业界尚无统一的标准。传统方法通常从整个研究区或人为划分的特定区域随机选取非滑坡样本,比如低坡度区域^[5]或距滑坡一定距离以外的区域^[6],这些方法所选取的非滑坡样本普遍存在准确性低、代表性差的问题。为此,一些学者利用数学统计模型无需负样本就能进行滑坡易发性区划的特性,开展易发性分区的初步划分,尝试从它划分的较低和极低易发区中获取更高质量的非滑坡样本。陈飞等^[7]从信息量模型划分的较低和极低易发区随机选取非滑坡点;在此基础上,周晓亭等^[8]加入从全区、低坡度区域选取非滑坡点作为对比;陈芯宇等^[9]采用确定性系数模型选取非滑坡点。他们的研究着重对比一种数学统计模型与传统方法所选取的负样本效果,并未探究不同数学统计模型在选取负样本方面的效果,找到选取非滑坡样本的最优方案。

针对机器学习模型负样本选取问题,本文以四川省汶川县、理县和茂县为研究区,将距滑坡点 1 km 以内的区域设为缓冲区,分别从信息量模型(information, I)、证据权模型(weight of evidence, WOE)、确定性系数模型(certainty factor, CF)和频率比模型(frequency ratio, FR)划分的滑坡较低和极低易发区以及缓冲区外(buffer, B)和整个研究区(globe, G)随机选取非滑坡样本,建立基于不同非滑坡样本选取方法的 SVM 模型(I-SVM、WOE-SVM、CF-SVM、FR-SVM、B-SVM、G-SVM,见表 1)进行滑坡易发性评价,旨在验证数学统计模型在非滑坡样本选取方面的有效性,找到选取非滑坡样本的最优方案,为当地地质灾害评估和预防治理提供依据。

1 研究区域概况及数据源

1.1 研究区概况

本文选取汶川县、理县和茂县为研究对象,研究区域位于四川省阿坝藏族羌族自治州东南部,地处四川盆地西北部边缘,总面积约 12 257.63 km²,属温带大陆性季风气候。研究区海拔 762 ~ 5 981 m,高差悬殊,地势起伏不平,西北高东南低,地形以高山峡谷为主,峰峦叠嶂,河谷深邃。地质构造复杂,地震活动频繁,进一步造成大量滑坡的发生。研究区地表水系为岷江和涪江,主要支流有岷江河、黑水河、杂谷脑河、草坡河和寿江等。研究区内历史滑坡点共有 176 个(图 1),为便于统计与计算,将研究区按照 30 m×30 m 的栅格单元进行划

表 1 基于不同非滑坡样本选取方法的 SVM 模型
Table 1 SVM model based on different non-landslide sample selection methods

模型名称	非滑坡样本选取方法: 基于数学统计模型(4种)				非滑坡样本选取方法: 常规(2种)	
	I-SVM	WOE-SVM	CF-SVM	FR-SVM	B-SVM	G-SVM
模型编号	1	2	3	4	5	6
命名规则	前面字母代表着选取非滑坡样本的方法,后面字母代表着所使用的SVM模型					
非滑坡样本选取区域	信息量模型划分的较低、极低易发区	证据权模型划分的较低、极低易发区	确定性系数模型划分的较低、极低易发区	频率比模型划分的较低、极低易发区	将距滑坡点1 km以内地区设为缓冲区,从缓冲区外随机选取	整个研究区

分, 共计 13 619 588 个栅格单元。

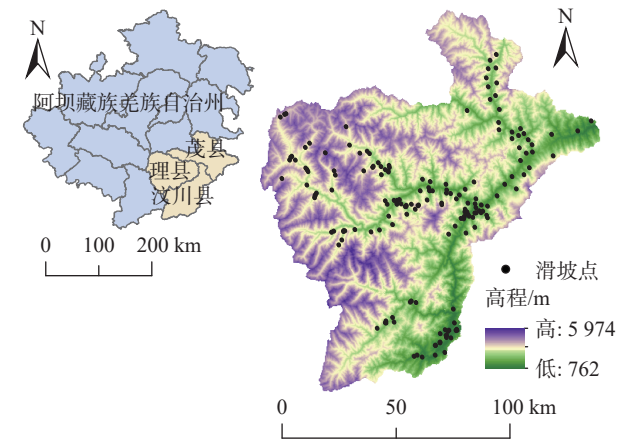


图 1 研究区滑坡分布图
Fig. 1 Landslide distribution map of the study area

1.2 数据源

本文研究采用的数据源主要: 1. 四川省自然资源厅和国家地球系统科学数据中心 (<http://www.geodata.cn>) 的 176 处滑坡点分布数据; 2. 地理空间数据云平台

(<http://www.gscloud.cn>) 30 m 分辨率的数字高程数据, 用于提取高程信息; 3. 全国地理信息资源目录服务系统 (<http://www.webmap.cn>) 的 1 : 25 万全国基础地理数据库的矢量水系数据和 30 m 全球地表覆盖数据; 4. 比例尺为 1 : 50 000 的区域地质构造图, 用于提取岩性因子信息。

2 研究方法

2.1 技术路线

技术路线见图 2, 过程详述如下:
步骤 1: 收集滑坡分布数据和评价因子数据, 采用皮尔逊相关系数和信息增益比筛选评价因子, 并采用 K-均值聚类法对评价因子进行分级。
步骤 2: 先采用 I、WOE、CF 和 FR 模型 (如表 1 所示, 模型编号 1—4) 进行初始滑坡易发性分区, 从上述模型划分的滑坡较低和极低易发区中随机选取 176 个非滑坡点; 对比试验将研究区距滑坡 1 km 以内的区域设为缓冲区, 分别从缓冲区外 (如表 1 所示, 模型编号

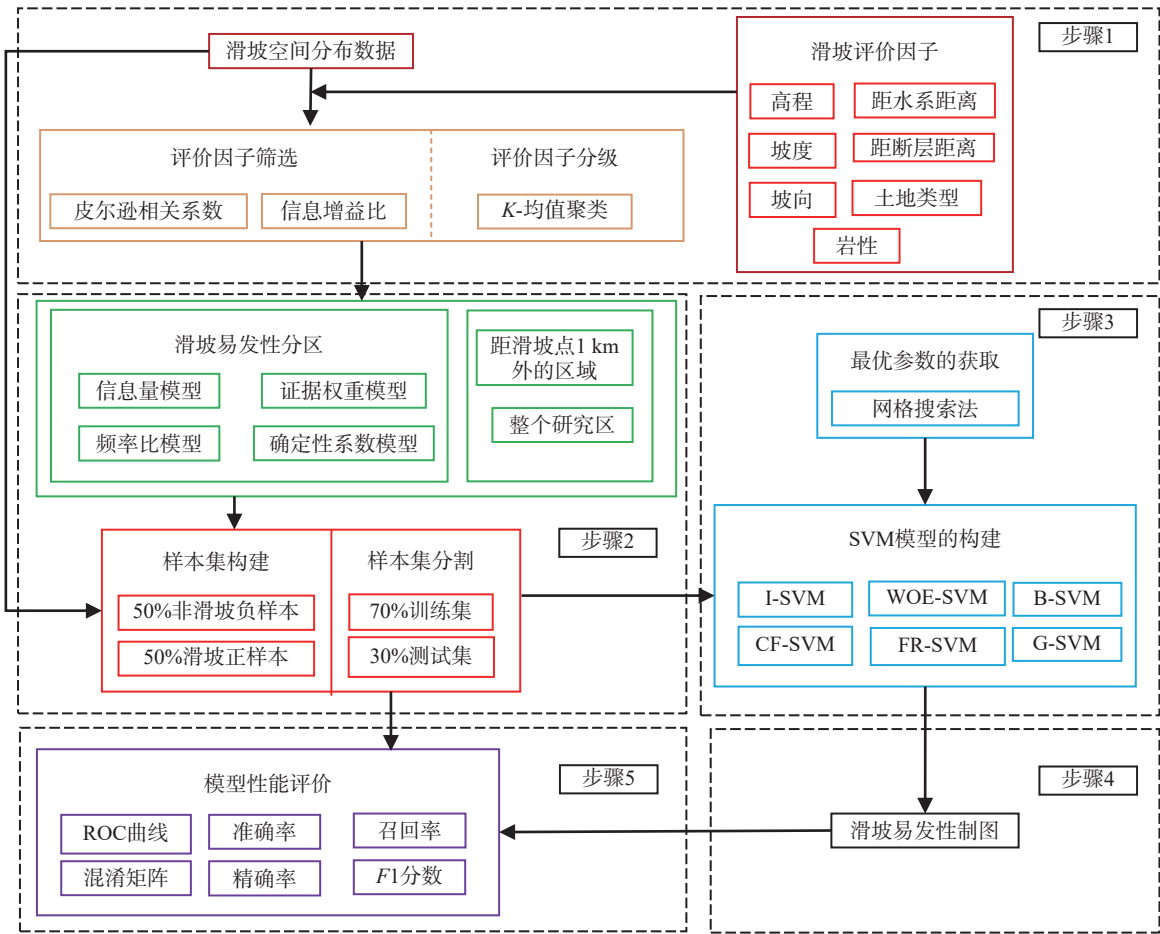


图 2 技术路线图
Fig. 2 Technology road mapping

5)和整个研究区(如表 1 所示,模型编号 6)随机选取 176 个非滑坡点,非滑坡点之间距离大于 1 km;将滑坡点赋值为 1,非滑坡点赋值为 0,作为模型输入参数;将 176 个滑坡点和选取的非滑坡点按 7:3 分为训练集和测试集。

步骤 3:使用训练集构建基于不同非滑坡样本选取方法的 SVM 模型,并采用网格搜索法获取模型的最优参数。

步骤 4:基于最优模型再次对研究区进行滑坡易发性制图。

步骤 5:使用测试集评价模型性能,验证数学统计模型在非滑坡样本选取方面的有效性,并找到选取非滑坡样本的最优方案。

2.2 SVM 模型

SVM 是一种基于统计学习理论的分类和回归方法,其基本思想是通过将输入向量映射到高维特征空间,构造一个最优的分类面或回归函数,实现对数据的学习和预测^[10]。

SVM 模型能够通过选取合适的核函数处理非线性和多分类问题,具有模型依赖的支持向量少、预测速度快等优点。在实际应用中,可通过寻优算法获得模型的最优参数,在提高模型精度的同时,提高模型的泛化程度。

2.3 基于数学模型预分区的 SVM 负样本的获取方法

2.3.1 I 模型

I 的理论基础源于概率论、信息论和工程类比法^[11],根据已知的滑坡分布资料和收集到的影响因子实测值计算出各个影响因子的信息量,将不同影响因子的数据转化为可以统一度量的信息量值^[12]。信息量模型计算公式如下:

$$I_i = \sum_{i=1}^n \ln \frac{A_i/A}{S_i/S} \quad (1)$$

式中: I_i ——评价单元总信息量;

A_i —— x_i 因子区域内发生滑坡面积;

A ——全区发生滑坡总面积;

S_i ——全区含有 x_i 所占面积;

S ——全区总面积。

其中,当信息量值大于 0 时,表示该影响因子对滑坡发生有促进作用;反之,该影响因子对滑坡发生有抑制作用。

2.3.2 WOE 模型

证据权模型的原理基于贝叶斯定理,它通过将历史

滑坡点与各种影响因子进行空间关联,计算得到影响因子在分类区间内的证据权值^[13]。证据权值的计算公式如下:

$$W_i^+ = \ln((A_i/A)/(B_i/B)) \quad (2)$$

$$W_i^- = \ln((a_i^+/A)/(a_i^-/B)) \quad (3)$$

$$W_{fi} = W_i^+ - W_i^- \quad (4)$$

式中: W_i^+ 、 W_i^- ——分别为 x_i 因子区域内的正相关、负相关权重值;

W_{fi} ——综合权重,表示 x_i 因子对滑坡变形失稳的权重;

B_i ——表示 x_i 因子内未发生滑坡面积;

B ——表示所有未发生滑坡的面积。

a_i^+ 、 a_i^- ——分别表示其他因子区域内发生滑坡面积、未发生滑坡面积;

2.3.3 CF 模型

CF 基于滑坡发生的概率函数,通过计算影响因子的确定系数来评估其对滑坡发生的贡献程度^[14],其计算公式如下:

$$CF = \begin{cases} \frac{PP_i - PP_s}{PP_s(1 - PP_i)} & (PP_i < PP_s) \\ \frac{PP_i - PP_s}{PP_i(1 - PP_s)} & (PP_i \geq PP_s) \end{cases} \quad (5)$$

$$PP_i = \frac{N_i}{S_i} \quad (6)$$

$$PP_s = \frac{N}{S} \quad (7)$$

式中: CF ——滑坡发生的确定性系数;

PP_i —— i 因子范围内滑坡数量与 i 因子所占面积比值;

PP_s ——滑坡总数与研究区总面积之比;

N_i —— x_i 因子区域内发生滑坡数量/处;

N ——全区发生滑坡总数量/处。

CF 值的取值范围为 $[-1, 1]$, CF 值越接近于 1,越容易发生滑坡; CF 值越接近于 0,越难发生滑坡; CF 值为 0 时,无法确定是否将发生滑坡。

2.3.4 FR 模型

频率比模型基于研究区域已发生滑坡点的分布和各影响因子之间的关系进行分析,提取不同等级下影响因子对滑坡发生的影响程度^[15]。

$$F_iR = \sum_{i=1}^n \frac{A_i/A}{S_i/S} \quad (8)$$

式中: $F_i R$ ——评价单元总频率比值。

频率比值大于 1, 表示该因子有利于滑坡的发生; 频率比值小于 1, 表明该因子不利于滑坡的发生。

3 评价因子的选取及影响因子分级

3.1 评价因子的选取

在评价因子的选取上, 周萍等^[16]选取高程、坡度、坡向、河网密度、距断层距离、降水量等因子对理县进行滑坡易发性评价; 付智勇等^[17]选取高程、坡度、坡向、剖面曲率、地层岩性和距断层距离分析汶川—映秀区域滑坡易发性。Wang 等^[18]的研究显示, 在汶川地区短时间内降雨量过大容易引发山体滑坡, 但并不意味着在任何情况降雨量越大越容易发生滑坡, 在植被覆盖度良好、地质条件稳定的高海拔地区, 即便年降雨量大于 807.88 mm, 滑坡依然很少发生。

基于上述分析, 本文选取坡度、坡向、高程、地层岩性、距断层距离、距水系距离和土地利用类型为研究区滑坡灾害易发性评价因子。考虑到研究区大部分为山区, 缺乏准确的降水资料, 故不考虑降雨因子对滑坡的影响。采用 K -均值聚类对评价因子分级, 结果见图 3。

(1) 坡度、坡向、高程

研究区内的地形因子包括坡度、坡向和高程坡度被分为了 9 类, 在 $24.6^\circ \sim 37.1^\circ$ 区间内的滑坡数量和面积最大; 坡向被分为了 5 类, $273^\circ \sim 360^\circ$ 范围内滑坡发生较多; 高程间接影响滑坡灾害的发育, $0 \sim 1\,704\text{ m}$ 范围内滑坡面积最大, 随着高程的增加, 滑坡面积逐渐减小, 见图 3(a)—(c)。

(2) 距水系距离

河流的存在会影响周围岩土体的物理和力学性质, 水流的持续冲刷进一步破坏了岩土体的稳定性, 导致滑坡的发生。研究区内滑坡点距水系距离最远为 1 972 m, 距水系的距离被分为 5 类, 滑坡大多发生于距水系 $0 \sim 314\text{ m}$ 范围内, 距水系距离越远, 滑坡数量和滑坡面积逐渐减小, 见图 3(d)。

(3) 距断层距离

构造断层是影响滑坡分布的主要因素之一。距断层距离被分为 8 类, 研究区内滑坡多发生于距断层 $0 \sim 1\,454\text{ m}$ 的范围内, 距断层越远, 滑坡面积整体呈下降趋势, 这表明断层对滑坡的发生有一定影响, 见图 3(e)。

(4) 岩性

岩性控制着滑坡的分布, 地层年代影响着岩石的分化程度。根据岩石坚硬程度, 本文将研究区岩性分为

5 类: 1. 坚硬岩石; 2. 较硬岩石; 3. 软硬相间岩石; 4. 较软岩石; 5. 软岩。分类结果见图 3(f)。

(5) 土地类型

根据全国地理信息资源目录服务系统提供的地表覆盖数据, 将研究区土地类型分为 4 类: 1. 耕地; 2. 林地; 3. 人类活动; 4. 其他用地。分类结果见图 3(g), 研究区大部分位于山区, 植被茂密, 人类活动面积较小, 滑坡多发生于耕地和林地。

3.2 因子间相关性分析

为了剔除冗余因子, 本文采用了皮尔逊相关系数检验因子相关性。皮尔逊相关系数^[19](Pearson correlation coefficient, PCC)是一种用于度量 2 个连续变量之间线性关系程度的统计量, 取值范围为 $[-1, 1]$ 。PCC 的绝对值越接近于 1 时, 表明两个变量间相关性越强。PCC 计算结果如图 4 所示, PCC 的绝对值均小于 0.5, 可认为评价因子间不存在显著相关性。

3.3 因子重要性分析

为了衡量评价因子对滑坡发生的影响程度, 本文采用信息增益比进行因子重要性分析。信息增益比^[20](information gain ratio, IGR)是用于特征选择的一种指标, 反映了影响因子对滑坡发生的重要程度, 计算结果如图 5 所示。7 个影响因子的 IGR 值均大于 0, 表明选取的影响因子对滑坡发生均有影响。其中, 土地类型对滑坡发生影响最大, 其次是坡向、距断层距离、岩性、距水系距离和高程, 坡度对滑坡发生影响最小。

综合考虑 PCC 和 IGR, 保留所有影响因子进行后续研究。

4 滑坡易发性评价

4.1 样本集的构建

本研究区滑坡绝大部分发生在坡度小于 55° 的地区, 当坡度大于 55° 时, 随着坡度的增加, 发生的滑坡越少, 因此不从低坡度区域选取非滑坡点。本文采取 6 种方式选取非滑坡样本: 使用 I、WOE、CF 和 FR 模型进行滑坡易发性分区(图 6), 再从他们划分的极低易发区和较低易发区中随机选取; 对比试验将距滑坡点 1 km 以内地区设为缓冲区, 分别从缓冲区外和整个研究区随机选取。

4.2 SVM 模型的构建

SVM 模型的构建, 是在 Python 环境下实现的^[21]。其中, SVM 模型采用了径向基函数(radial basis function, RBF)作为核函数, 该模型性能受参数 Gamma 和惩罚因子(C)影响很大。Gamma 控制径向基函数核的大

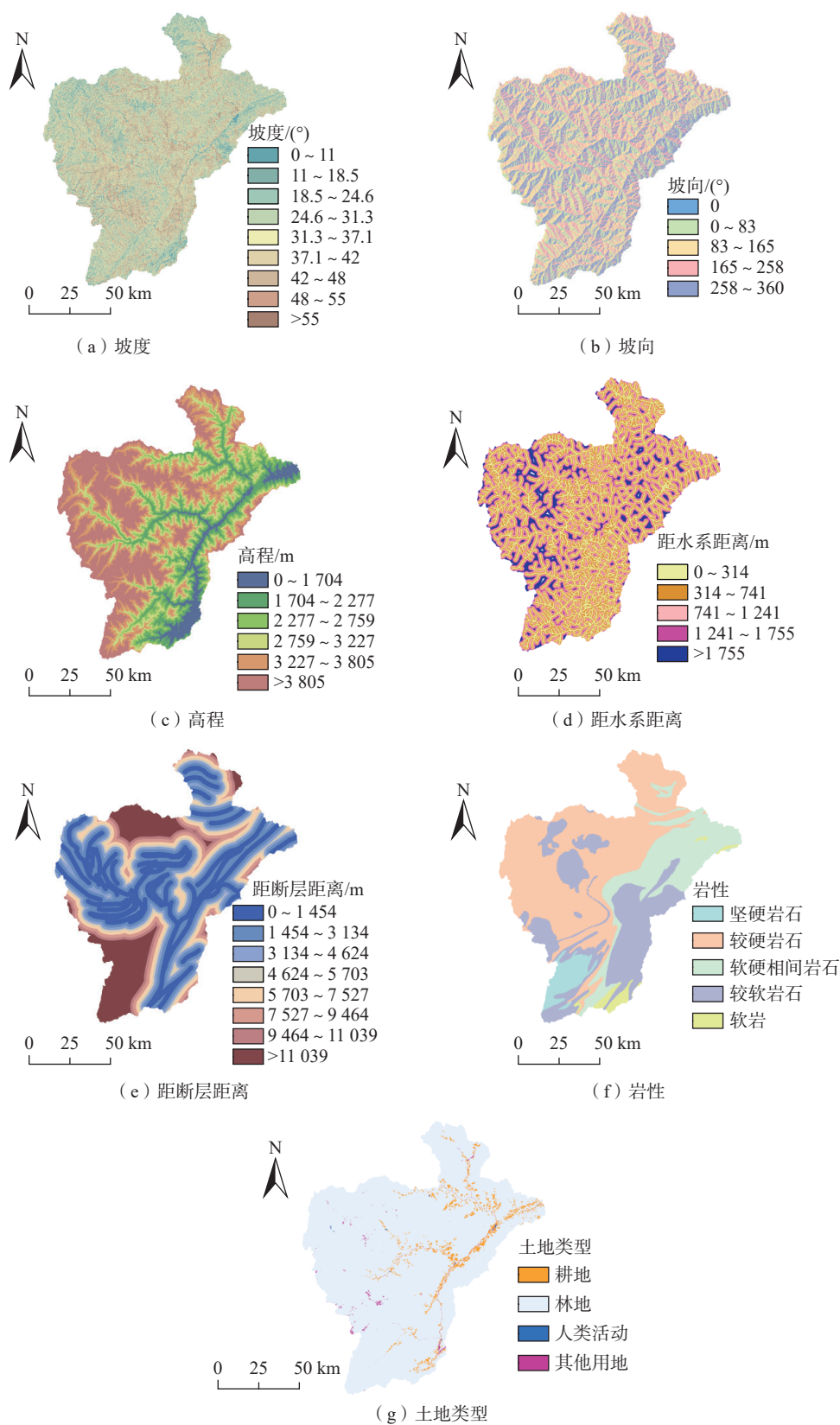


图 3 环境因子分级

Fig. 3 Maps of the landslide conditioning factors

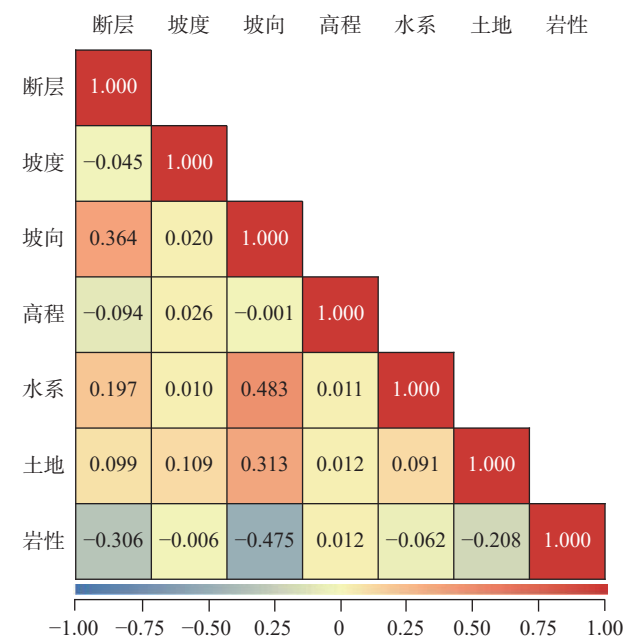


图 4 皮尔逊相关系数

Fig. 4 Pearson correlation coefficients

小, C 控制边界线的硬度。为了提高模型的精度, 减少过拟合, 本文采取网格搜索法获取了各个 SVM 模型的最优参数, 如表 2 所示。

将研究区输入已训练好的 SVM 模型, 获得了研究

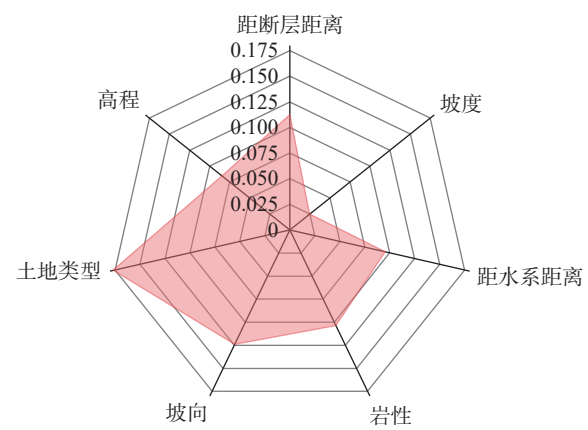


图 5 信息增益比

Fig. 5 Information gain ratio

区的滑坡易发性指数。利用自然断点法对模型计算出的滑坡易发性指数进行分级, 将研究区划分为极低、较低、中等、较高、极高易发区 5 个等级, 结果如图 7 所示。各个模型预测的滑坡易发性分级面积和滑坡面积占比如表 3 所示。由图 7 和表 3 可知: I-SVM 和 WOE-SVM 预测趋势基本吻合, 能够精确地划分滑坡易发区; CF-SVM 在部分区域识别表现不佳, 比如研究区的东南区域; FR-SVM 对极高易发区识别效果较差, 存在将非滑坡区域识别为潜在滑坡区域情况; B-SVM 和 G-SVM

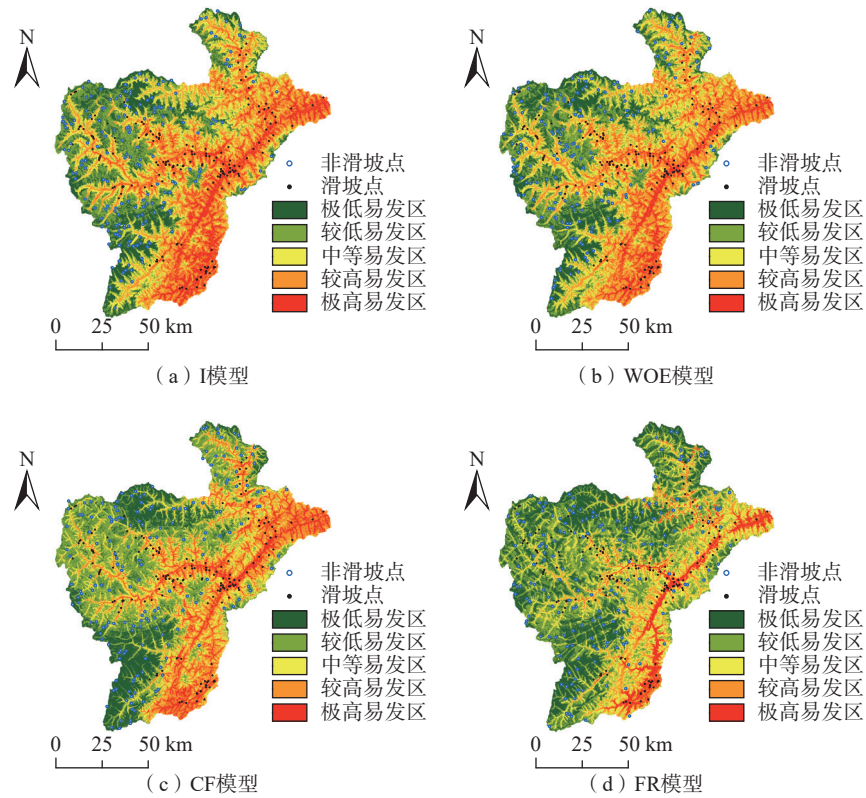


图 6 不同模型的滑坡易发性分区

Fig. 6 Landslide susceptibility zoning by different models

表 2 不同模型的最优参数

Table 2 Optimal parameters for different models

参数	I-SVM	WOE-SVM	CF-SVM	FR-SVM	B-SVM	G-SVM
Gamma	0.01	0.01	0.02	0.1	0.01	0.02
C	160	180	8	2	40	1

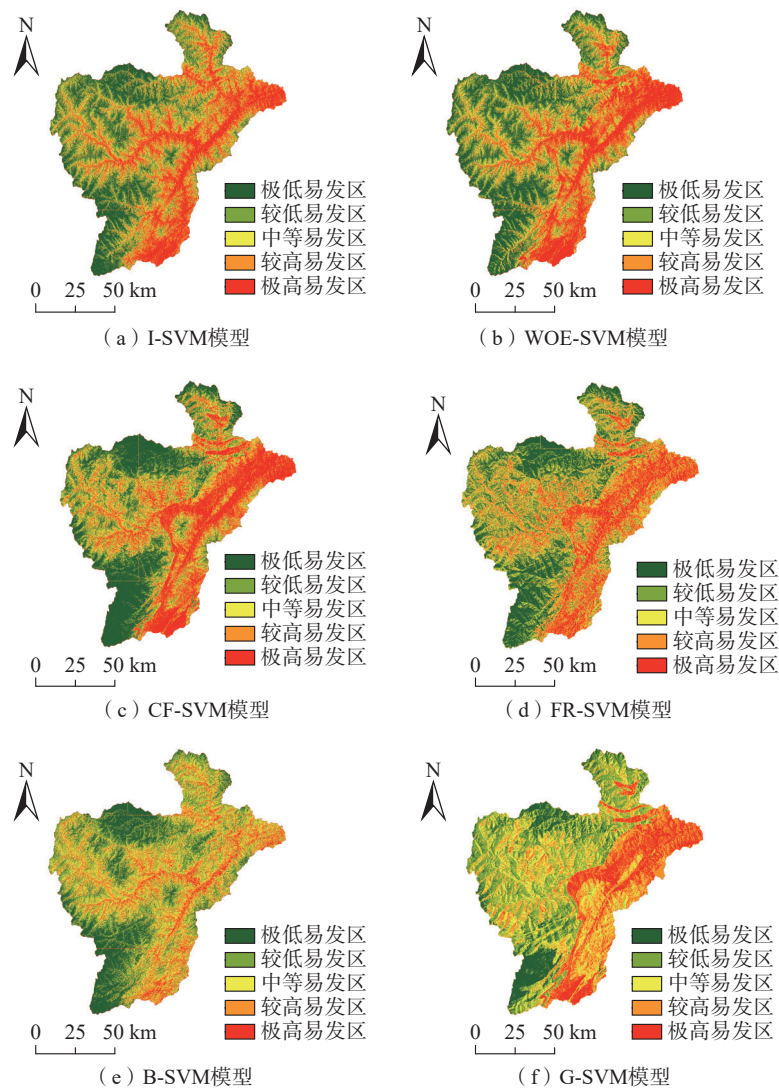


图 7 基于 SVM 模型的滑坡易发性分区

Fig. 7 Landslide susceptibility zoning generated based on SVM model

均存在将潜在滑坡区域识别为非滑坡区域情况, 预测效果较差。

B-SVM 划分的极高易发区面积为 1 104.34 km², 对应的滑坡面积占比为 31.45%, 低于其他模型。这是由

表 3 易发性分级面积及滑坡面积占比

Table 3 Classification area of susceptibility and proportion of landslide area

易发性分级		极低	较低	中等	较高	极高
易发性 分级面积/km ²	I-SVM	2 696.75	2 244.36	2 006.25	2 189.49	3 120.79
	WOE-SVM	3 058.40	2 250.93	1 869.58	1 984.89	3 093.82
	CF-SVM	3 190.05	1 975.34	1 817.33	2 060.21	3 214.70
	FR-SVM	2 780.17	2 293.44	2 265.99	2 454.95	2 463.09
	B-SVM	2 647.25	3 111.23	3 095.24	2 299.57	1 104.34
	G-SVM	1 381.98	3 186.95	3 089.52	2 652.01	1 947.17

(续表)

易发性分级		极低	较低	中等	较高	极高
滑坡面积 占比/%	I-SVM	0.11	0.59	4.06	13.15	82.10
	WOE-SVM	1.03	1.19	4.98	20.78	72.01
	CF-SVM	0.10	5.41	4.75	19.55	70.19
	FR-SVM	0.21	2.56	15.66	10.94	70.64
	B-SVM	1.58	5.85	18.14	42.98	31.45
	G-SVM	3.31	11.46	25.82	13.97	45.44

于从缓冲区外选取非滑坡点仅考虑到了滑坡空间分布情况,没有考虑到非滑坡区域的特征。但实际工作中无法完全统计研究区内所有滑坡区域与潜在滑坡区域,且缓冲区的设置依赖于人工判断和专家经验,不可避免地存在从潜在滑坡区域选取非滑坡点,导致 B-SVM 预测更偏向于非滑坡区域,而对滑坡高易发区识别效果不佳。在综合考虑较高和极高易发区的情况下, B-SVM 的滑坡面积占比达到了 74.43%, 优于 G-SVM 的 59.41%, 说明从缓冲区外随机选取非滑坡点在一定程度上提高了非滑坡样本的准确性,从而提高了模型的预测精度。

4.3 模型精度评价

为了检验评价结果的准确性和合理性,采用测试数据集评价模型性能,并选取 ROC 曲线、混淆矩阵、准确性、精确率、召回率和 F1 分数等 6 个参数作为评价指标。图 8 为各个模型的 ROC 曲线,图 9 展示了各个模型使用测试数据集的性能。由图 8、图 9 可知: 6 种模

型预测精度由高到低依次为: I-SVM>WOE-SVM>CF-SVM>FR-SVM>B-SVM>G-SVM。

G-SVM 是从未发生滑坡区域随机选取非滑坡样

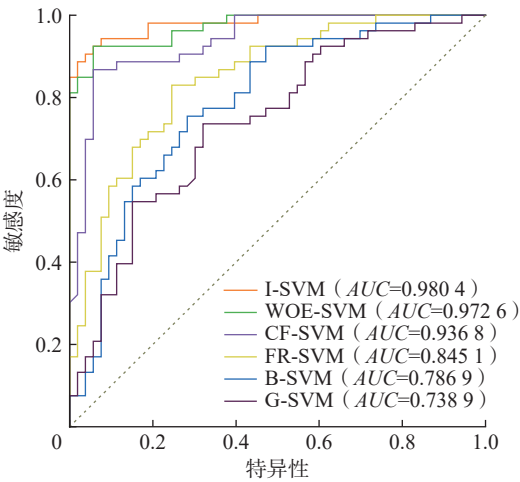


图 8 不同模型的 ROC 曲线

Fig. 8 ROC curve of different models

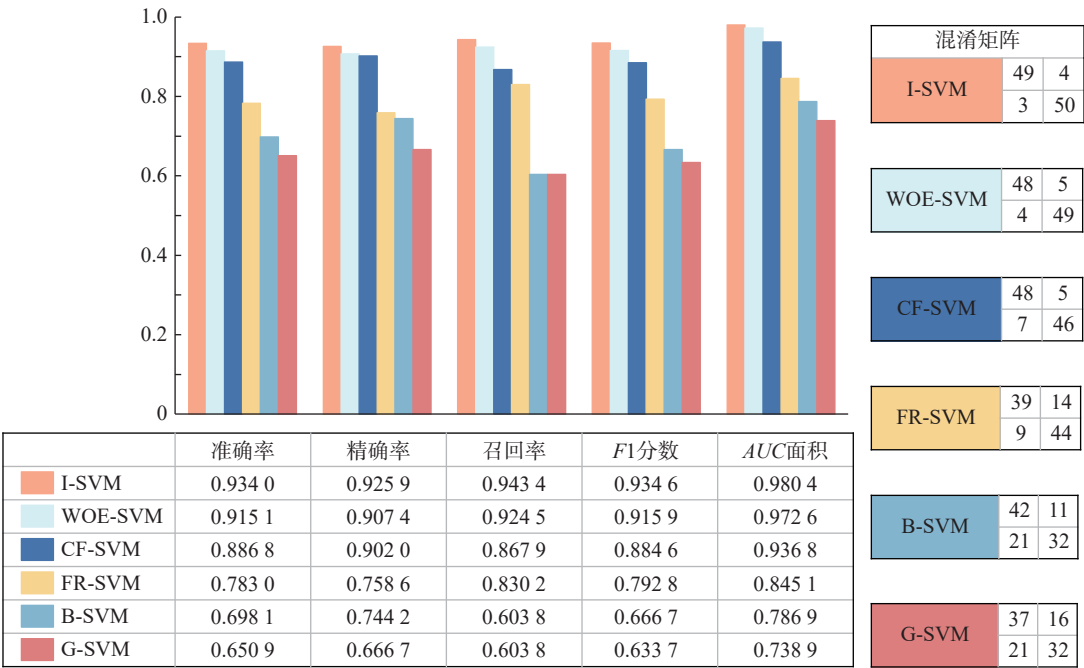


图 9 模型性能评价指标

Fig. 9 Evaluation index of the model performance

本,这种方法避免了人为选择干扰,其正负样本预测效果也比较均匀,但该方法存在一定随机性,不可避免地在部分地质条件下过多或过少地选择了非滑坡点,使所选样本在空间上不稳定,不能很好地代表整个研究区的非滑坡点特征,导致模型精度下降。

数学统计模型在进行滑坡易发性评价时考虑了滑坡在各个因子的不同分级下的分布情况,其划分的滑坡较低、极低易发区更好地反映了非滑坡区域的特征,所选取的负样本准确性更高,因此 I-SVM、WOE-SVM、CF-SVM、FR-SVM 模型精度和预测结果均优于 B-SVM 和 G-SVM。其中, I-SVM 精度最高,信息量模型是选取非滑坡样本的最优方案。

在操作难度方面,从整个研究区随机选取非滑坡样本是最容易的;从缓冲区外选取非滑坡点也比较方便,但依赖于专家经验确定缓冲区范围;与前两种方法相比,使用数学统计模型预分区选取非滑坡点效果最好,但需要先对研究区进行滑坡易发性预分区,增加了很多工作量。

5 结论

(1)相比于传统方法从整个研究区或人为划分的区域选取负样本,采用数学统计模型选取的负样本准确性更高,其中信息量模型是选取非滑坡样本的最优方案。

(2)6 种模型(I-SVM、WOE-SVM、CF-SVM、FR-SVM、B-SVM、G-SVM)的滑坡易发性分区图表现出一定的相似性,研究区极高和较高易发区主要分布在高程较低的山谷地区,多为耕地和林地区域,越接近河流和断层发生滑坡风险越大。

参考文献(References):

- [1] XIA Ding, TANG Huiming, SUN Sixuan, et al. Landslide susceptibility mapping based on the germinal center optimization algorithm and support vector classification [J]. *Remote Sensing*, 2022, 14(11): 2707.
- [2] 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)]
- [3] DENG Hui, WU Xiantan, ZHANG Wenjiang, et al. Slope-unit scale landslide susceptibility mapping based on the random forest model in deep valley areas [J]. *Remote Sensing*, 2022, 14(17): 4245.
- [4] QUAN Hechun, JIN Guangri, DESTTECH PUBLICAT I. Landslide susceptibility mapping in Changbai Mountain Area using GIS and artificial neural network(ANN); proceedings of the international conference on GIS and resource management (ICGRM), Guangzhou, PEOPLES R CHINA, F Jan 03-05, 2014 [C]. 2014.
- [5] 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.
- [6] DOU Jie, YUNUS A P, BUI D T, et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan [J]. *The Science of the Total Environment*, 2019, 662: 332-346.
- [7] 陈飞,蔡超,李小双,等.基于信息量与神经网络模型的滑坡易发性评价[J].岩石力学与工程学报,2020,39(增刊1):2859-2870.[CHEN Fei, CAI Chao, LI Xiaoshuang, et al. Evaluation of landslide susceptibility based on information quantity and neural network model [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2020, 39(Sup 1): 2859-2870. (in Chinese)]
- [8] 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 selected by coupling information value method [J]. *Advanced Engineering Sciences*, 2022, 54(3): 25-35. (in Chinese with English abstract)]
- [9] 陈芯宇,师芸,温永啸,等.基于确定性系数与支持向量机的滑坡易发性评价[J].科学技术与工程,2023,23(2):518-527.[CHEN Xinyu, SHI Yun, WEN Yongxiao, et al. Landslide susceptibility evaluation based on certainty factor and support vector machines [J]. *Science Technology and Engineering*, 2023, 23(2): 518-527. (in Chinese with English abstract)]
- [10] LIU Jiping, LIN Rongfu, XU Shenghua, et al. Landslide susceptibility evaluation based on optimized support vector machine [J]. *Proceedings of the ICA*, 2021, 4: 1-4.
- [11] 周天伦,曾超,范晨,等.基于快速聚类-信息量模型的汶川及周边两县滑坡易发性评价[J].中国地质灾害与防治学报,2021,32(5):137-150.[ZHOU Tianlun, ZENG Chao, FAN Chen, et al. Landslide susceptibility assessment based on K-means cluster information model in Wenchuan and two neighboring counties, China [J]. *The Chinese Journal of*

- Geological Hazard and Control, 2021, 32(5): 137 – 150. (in Chinese with English abstract)]
- [12] 熊小辉,汪长林,白永健,等.基于不同耦合模型的县域滑坡易发性评价对比分析——以四川普格县为例 [J]. 中国地质灾害与防治学报, 2022, 33(4): 114 – 124. [XIONG Xiaohui, WANG Changlin, BAI Yongjian, et al. Comparison of landslide susceptibility assessment based on multiple hybrid models at county level: A case study for Puge County, Sichuan Province [J]. The Chinese Journal of Geological Hazard and Control, 2022, 33(4): 114 – 124. (in Chinese with English abstract)]
- [13] NEUHÄUSER B, DAMM B, TERHORST B. GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model [J]. *Landslides*, 2012, 9(4): 511 – 528.
- [14] 陈建国,钟连祥.基于CF-AHP耦合模型的高速公路沿线崩塌地质灾害易发性评价——以蓉遵高速公路土城——旺隆段为例 [J]. 中国地质灾害与防治学报, 2023, 34(6): 105 – 115. [CHEN Jianguo, ZHONG Lianxiang. Assessment of rockfall susceptibility along the expressway based on the CF-AHP coupling model: A case study of the Tucheng-Wanglong section of the Rongzun expressway [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(6): 105 – 115. (in Chinese with English abstract)]
- [15] BERHANE G, KEBEDE M, ALFARAH N, et al. Landslide susceptibility zonation mapping using GIS-based frequency ratio model with multi-class spatial data-sets in the Adwa-Adigrat Mountain chains, northern Ethiopia [J]. *Journal of African Earth Sciences*, 2020, 164: 103795.
- [16] 周萍,邓辉,张文江,等.基于信息量模型和机器学习方法的滑坡易发性评价研究——以四川理县为例 [J]. 地理科学, 2022, 42(9): 1665 – 1675. [ZHOU Ping, DENG Hui, ZHANG Wenjiang, et al. Landslide susceptibility evaluation based on information value model and machine learning method: A case study of Lixian County, Sichuan Province [J]. *Scientia Geographica Sinica*, 2022, 42(9): 1665 – 1675. (in Chinese with English abstract)]
- [17] 付智勇,李典庆,王顺,等.基于多时空滑坡编录和TrAdaBoost迁移学习的滑坡易发性评价 [J]. 地球科学, 2023, 48(5): 1935 – 1947. [FU Zhiyong, LI Dianqing, WANG Shun, et al. Landslide susceptibility assessment based on multitemporal landslide inventories and TrAdaBoost transfer learning [J]. *Earth Science*, 2023, 48(5): 1935 – 1947. (in Chinese with English abstract)]
- [18] WANG Xiao, LI Shaoda, LIU Hanhu, et al. Landslide susceptibility assessment in Wenchuan County after the 5•12 magnitude earthquake [J]. *Bulletin of Engineering Geology and the Environment*, 2021, 80(7): 5369 – 5390.
- [19] 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)]
- [20] 姚皖路,赵俊三,李坤.基于信息增益比-支持向量机的泥石流易发性评价 [J]. *城市勘测*, 2023(5): 181 – 186. [YAO Wanlu, ZHAO Junsan, LI Kun. Assessment of debris flow susceptibility based on information gain ratio-support vector machine [J]. *Urban Geotechnical Investigation & Surveying*, 2023(5): 181 – 186. (in Chinese with English abstract)]
- [21] HUANG Wubiao, DING Mingtao, LI Zhenhong, et al. An efficient user-friendly integration tool for landslide susceptibility mapping based on support vector machines: SVM-LSM toolbox [J]. *Remote Sensing*, 2022, 14(14): 3408.