

doi:10.3969/j.issn.2097-0013.2025.01.015

基于多种评价模型在斜坡地质灾害易发性评价中的 对比分析——以武汉市为例

谢扬龙^{1,2}, 齐 信^{1,3*}, 高 鹏^{1,2}, 王英豪³

XIE Yang-Long^{1,2}, QI Xin^{1,3*}, GAO Peng^{1,2}, WANG Ying-Hao³

1. 中国地质调查局武汉地质调查中心(中南地质科技创新中心), 湖北 武汉 430205; 2. 中国地质大学(武汉)地质调查
研究院, 湖北 武汉 430074; 3. 中国地质大学(武汉)工程学院, 湖北 武汉 430074

1. Wuhan Center, China Geological Survey (Geosciences Innovation Center of Central South China), Wuhan 430205, Hubei, China;

2. Institute of Geological Survey, China University of Geoscience, Wuhan 430074, Hubei, China;

3. Faculty of Engineering, China University of Geoscience, Wuhan 430074, Hubei, China

摘要: 地质灾害易发性评价是防灾减灾工作中不可或缺的部分, 挑选行之有效的评价方法与评价模型进行地质灾害易发性评价具有重要意义。本文以武汉市斜坡地质灾害(崩塌和滑坡)为研究对象, 选取坡度、坡向、高程、工程地质岩组、植被覆盖率等 9 项评价因子, 分别使用信息量法(IV)、确定系数法(CF)及随机森林模型(RF)进行地质灾害易发性评价, 然后使用受试者工作特异性曲线(ROC)检验评价模型精确度并对比三种模型的评价结果。结果表明: (1) 三种模型均能正确反映研究区斜坡地质灾害的发育特征, 高易发区与极高易发区集中在研究区北部、中西部的高山地区及经河流剥蚀夷平形成的剥蚀堆积丘陵区域, 低易发区占绝大部分, 主要分布于长江沿岸和湖泊周边的平原地带及绝大部分低山平原地区。(2) 三种模型 AUC 面积从高到低依次为 IV>CF>RF, 其中 IV 和 CF 模型得到的区划结果具有较高的相似度, 同时二者的极高易发区的灾害点密度均要高于 RF 模型。(3) 坡度、高程、工程地质岩组、斜坡结构类型是评价体系中重要性较高的因子, 说明研究区斜坡地质灾害的发育因素主要与地形地貌、地层岩性有关。

关键词: 斜坡地质灾害易发性评价; 评价因子; 信息量法; 确定系数法; 随机森林; 武汉市

中图分类号: P694

文献标识码: A

文章编号: 2097-0013(2025)-01-0185-12

Xie Y L, Qi X, Gao P and Wang Y H. 2025. Comparative Analysis of Slope Geological Hazard Susceptibility Assessment Based on Multiple Evaluation Models: A Case Study of Wuhan, China. *South China Geology*, 41(1): 185–196.

Abstract: The susceptibility evaluation of geological disasters is an indispensable part of disaster prevention and mitigation. It is of great significance to select effective evaluation methods and models for the susceptibility evaluation of geological disasters. This paper takes the slope geological disasters (collapse and landslide) in Wuhan city as the research object, selects 9 evaluation factors such as slope, slope direction, elevation, engineering geological rock group and vegetation coverage, and uses the information method (IV), the determination coefficient method (CF) and the random forest model (RF) to evaluate the susceptibility of

收稿日期: 2024-11-07; 修回日期: 2024-12-04

基金项目: 中国地质调查局项目(DD20221734)

第一作者: 谢扬龙(2000—), 男, 硕士研究生, 主要从事灾害地质、工程地质的调查与研究工作, E-mail: 2295708640@qq.com

通讯作者: 齐 信(1983—), 男, 正高级工程师, 主要从事灾害地质、工程地质、环境地质的调查与研究工作, E-mail: qx_cdu@126.com

geological disasters, and then uses the receiver operating characteristic curve (ROC) to test the accuracy of the evaluation model and compare the evaluation results of the three models. The results show that: (1) The three models can correctly reflect the development characteristics of slope geological disasters in the study area. The high-prone areas and extremely high-prone areas are concentrated in the mountainous areas in the northern, central and western parts of the study area and the eroded accumulation hilly areas formed by river denudation and leveling. The low-prone areas account for the vast majority and are mainly distributed in the plain areas around the banks and lakes of the Yangtze River and most of the low-mountain plain areas. (2) The AUC area of the three models from high to low is $IV > CF > RF$. The zoning results obtained by the IV and CF models have high similarity, and the disaster point density of the extremely high-prone areas of the two models is higher than that of the RF model. (3) Slope, elevation, engineering geological rock group and slope structure type are the most important factors in the evaluation system, indicating that the development factors of slope geological disasters in the study area are mainly related to topography and lithology.

Key words: susceptibility evaluation of slope geological disasters; evaluation factor; information quantity method; determination of coefficient method; random forest; Wuhan city

我国地处欧亚大陆东部,地形地貌多样,地质条件复杂,加上季风气候的影响,使得崩塌、滑坡、泥石流等地质灾害频繁发生。开展地质灾害易发性评价有助于了解地质环境条件对地质灾害发生的影响,为政府和相关部门提供决策依据,以便采取有效的预防和减灾措施(彭建兵等, 2022)。通过对特定地区地质灾害的易发性进行评估,可以明确该地区的重点保护对象和需要采取的防治措施,有助于实现资源、环境和经济的协调发展。

目前,国内外学者针对地质灾害易发性评价的研究方法,主要分为经验模型(专家打分法、层次分析法等)、统计模型(信息量法、确定系数法等)和机器学习模型(支持向量机、随机森林模型、人工神经网络模型等)等(刘坚等, 2018; 连志鹏等, 2022; 王雪冬等, 2022; 陈建国和钟连祥, 2023; 翟文华等, 2023; 张威等, 2023; 章昱等, 2023)。孙剑锋等(2023)利用灰色关联度与层次分析耦合的方法建立 3 种评价模型,分别分析了 9 项评价因子与历史灾害点的关联度并对结果进行了可靠性分析。于喜坤等(2023)基于信息量模型与确定系数模型计算灾害点的信息量与确定系数并以此为依据完成易发性制图。孟凡奇等(2023)利用 GIS 平台,运用信息量模型和层次分析法进行地质灾害易发性分区。邵大江等(2023)以云南省南华县 2015 年地质灾害数据为基础,选

择 11 个相关因子,基于均值法,采取梯度提升树算法(XGBoost、LightGBM、CatBoost)、信息量模型与地理加权回归模型开展了地质灾害易发性评价。研究表明:地理加权回归模型预测结果存在过拟合现象,信息量模型则存在欠拟合现象。黄敏(2023)选取支持向量机模型、随机森林模型和人工神经网络模型进行地质灾害易发性分析,研究表明:支持向量机模型的频度比分析值更具梯度,是最适合的易发性分析模型。Zhang S H et al.(2024)基于两种评价单元与四种评价模型,选取海拔、地形起伏等 9 个灾害因子作为评价指标,借助 ArcGIS 平台进行地质灾害易发性评价,研究表明四种评价模型中斜坡评价单元的精确度高于栅格评价单元,其中斜坡单元与随机森林模型的集成达到了最高的预测精度。Sun K et al.(2024)选取青藏高原南部公路为研究对象,采用支持向量机(SVM)模型并使用贝叶斯算法进行超参数优化,选取高程、坡度、坡向等 9 个影响因子进行崩塌灾害的易发性评价,研究发现:训练数据集和测试数据集的准确率分别为 99.15% 和 96.32%;岩性和高程因子是影响性较大的因子。Wang Z P et al.(2024)采用随机森林(RF)模型和频率比随机森林(FR-RF)模型,创新性地探究了不同缓冲距离区间内的非灾害样本点对易发性评价结果的影响,结果表明:评价的精度随着距

离的增大呈现出先升高后降低的趋势,并且同一距离区间内 RF 模型的精度高于 FR-RF 模型。Li R W et al.(2024)利用随机森林(RF)赋能信息量的方法进行地质灾害易发性评价,该方法基于参数优化的随机森林模型计算目标权重,将这些权重分配给每个评估因子,然后对信息进行加权叠加。实验结果表明,该方法的准确度和可靠性优于传统方法。

上述三种模型中经验模型简化了地质灾害孕灾因子的影响机制,主要通过人为确定权重反映其对灾害发生的贡献率,并不能大面积推广使用,具有一定的局限性;统计模型的主要流程是先选定评价因子,后将评价因子分级并统计各评价因子分级区间的面积及灾害点的密度,最后计算相关指标并进行易发性评价,这种模型忽略了评价因子自身的特异性,且不同评价因子的选择也会影响评价结果的准确性;机器学习模型是通过数学理论进行易发性评价,该方法对于高维度数据的泛化能力强,有较强的抗拟合性,不足之处是模型的具体参数选择较为困难,在样本数据不充足的情况下,预测精度较低(Pradhan, 2013; Reichenbach et al., 2018; Lombardo et al., 2020; 焦伟之等, 2022; Jiang N et al., 2024; Sun K et al., 2024)。总之,不同研究区域的地质条件、气候因素存在差异,在不同精度的比例尺研究中,模型的适用性也不同。因此,挑选行之有效的评价方法与评价模型进行地质灾害易发性评价具有重要意义。

综上所述,地质灾害易发性评价是一项多元化、复杂化的工作。本文以湖北省武汉市为研究区,选取坡度、坡向、高程、工程地质岩组等 9 项评价因子,分别运用信息量法、确定系数法及随机森林模型进行地质灾害易发性评价与对比研究,探究不同评价方法的有效性和评价模型的可靠性,为武汉市及其他地区的地质灾害易发性评价与区划防控提供科技支撑与经验借鉴。

1 研究区概况和地质灾害形成机理

1.1 研究区概况

武汉市地处江汉平原东部、长江中游,三面环

山的半封闭型盆地中心,中间低平,南北丘陵、岗垄环抱,北部低山林立。地理位置为东经 113°41′~115°05′、北纬 29°58′~31°22′,总面积 8 569.15 km²,划分为 15 个行政区。全市海拔高度在 19.2 m 至 873.7 m 之间,大部分在 50 m 以下,最高峰为北部黄陂区与孝感市交界处的双峰尖,海拔为 873.7 m(图 1)。研究区地处北温带,属中亚热带过渡的湿润季风气候,多年平均气温在 17℃左右,平均降水量 1 233.3 mm,最大全年降雨量为 1 487.5 mm。研究区地质构造以新华夏构造体系为主,地貌属鄂东南丘陵经汉江平原东缘向大别山南麓低山丘陵过渡地区。研究区目前共发现 66 处滑坡灾害点,33 处崩塌灾害点,已经严重危害了当地居民的生命健康及财产安全。因此,开展武汉市的斜坡地质灾害易发性评价显得尤为重要。

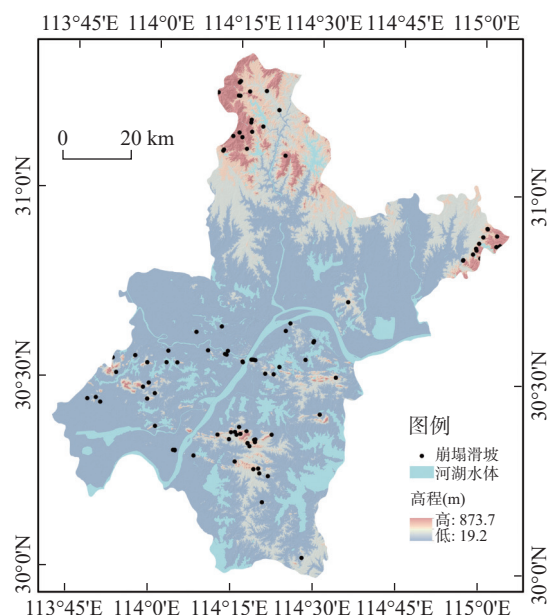


图 1 武汉市高程与地质灾害分布概况图

Fig. 1 Overview map of elevation and geological hazard distribution in Wuhan city

1.2 崩塌与滑坡形成机理

当前研究区域内共记录到 33 个崩塌点,包括 31 个小型和 2 个中型崩塌,无大型崩塌。研究区的崩塌形成与岩石内部的结构面类型有很大关系:控制崩塌的主要结构面为节理裂隙面,占崩塌总数的 83%。这些崩塌多发生在硬质至较硬质的片岩和中至厚层碎屑岩中,岩石中存在的易软化夹层经过长时间的强烈风化作用,使得节理裂隙

面更为发育。其次,崩塌主要发生于低山丘陵地带,该区岩石以石英砂岩、粉砂岩、片岩和片麻岩为主,这些岩石容易发生脆性破坏,在经历了构造和卸荷裂隙的影响后,岩石破碎性增加。经过长期的风化和侵蚀作用,加之重力的影响,最终导致崩塌。

目前研究区内共识别出 66 例滑坡,包括 57 例土质滑坡、7 例岩质滑坡和 2 例岩土混合滑坡。这些滑坡的分布与崩塌点分布位置大致吻合,主要位于低矮的山丘和丘陵地形区域。土质滑坡的成因可以归结为以下两个方面:(1)受到土壤结构松散、高孔隙度和良好的渗透性影响,结合底层基岩较低的透水性,难以形成有效的阻水层,地表水的渗透导致土层饱和,地下水的下渗作用软化土层,增加土层重量及孔隙水压,从而降低土层的力学稳定性,易在土层内部或与基岩接触面上形成潜在的滑移带。(2)研究区城市化速度的加快,大规模的土地开发活动改变了原有的应力分布,破坏了斜坡的自然应力平衡状态,在缺乏有效的坡面防护和排水措施的情况下,持续降雨会沿着土壤的裂缝或孔隙渗透,增加坡体的含水量和自重,进而提高了滑动力。此外,原生裂隙的存在以及干湿循环等自然风化作用促进了坡面和坡顶裂隙的发展,降雨通过这些通道增加了坡体的静水压力,增大了向下的滑动力。

2 数据来源和评价方法

2.1 数据来源

本研究区数据选择坡度、坡向、高程、工程地质岩组、植被覆盖率、斜坡结构类型、距构造距离、距道路距离、距水系距离共 9 个地质灾害影响因子,构建研究区斜坡地质灾害评价因子数据,数据来源如表 1 所示。

2.2 信息量法 (IV)

信息量模型是利用信息的概念来分析在各种因素综合作用下滑坡易发性程度的模型,该模型简便实用,能正确反映地质灾害的基本规律,便于推广应用。其基本观点认为,斜坡地质灾害的产生与否或发生概率与预测过程中所获取的已变形或破坏区域的信息数量和质量有关,信息量值越大说明该评价因子对灾害发生的贡献程度越大(温鑫等, 2022)。具体计算公式如下:

$$I(x_i, H) = \ln \frac{P(x_i|H)}{P(x_i)} = \ln \frac{N_i/N}{S_i/S} \tag{1}$$

式中: $I(x_i, H)$ 为斜坡地质灾害发生时评价因子计算的信息量, N_i 为研究区内因子划分区间中斜坡地质灾害数量; N 为研究区内斜坡地质灾害总数; S_i 为研究区内因子划分区间中评价栅格数量; S 为研究区内评价栅格总数。

表 1 研究区数据源
Table 1 Data sources for the study area

数据名称	数据来源	数据处理方式
坡度	DEM	ArcGIS空间分析工具提取坡度
坡向	DEM	ArcGIS空间分析工具提取坡向
高程	NASA地球科学数据网站(https://nasadaacs.eos.nasa.gov/)	下载12.5 m精度DEM, 采用二次曲面函数对高程异常进行拟合校正
工程地质岩组	湖北省地质调查院、湖北省地质环境总站、湖北省武汉水文地质工程地质大队等单位完成的基础调查数据	ArcGIS重分类工具矢量转栅格
植被覆盖率	地理空间数据云	下载Landsat 8 OLI 30 m精度影像 ENVI5.3处理分析
斜坡结构类型	湖北省地质调查院、湖北省地质环境总站、湖北省武汉水文地质工程地质大队等单位完成的基础地质调查数据	ArcGIS重分类工具矢量转栅格
距构造距离	湖北省地质调查院、湖北省地质环境总站、湖北省武汉水文地质工程地质大队等单位完成的基础地质调查数据	ArcGIS空间分析工具缓冲区分析
距道路距离	全国地理信息资源目录服务系统	ArcGIS空间分析工具缓冲区分析
距水系距离	全国地理信息资源目录服务系统	ArcGIS空间分析工具缓冲区分析

2.3 确定系数法 (CF)

确定系数值表示一个随机变量与多个随机变量关系的数字特征,其取值范围在-1 至 1 之间(赵晓燕等, 2022)。在地质灾害易发性评价中,假设过去发生地质灾害的条件和未来发生地质灾害的条件相同,具体计算公式如下:

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)} & PP_a \geq PP_s \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)} & PP_a < PP_s \end{cases} \quad (2)$$

式中: PP_a 为评价因子 a 划分区间中斜坡地质灾害数量与评价因子 a 划分区间所占面积的比值; PP_s 为研究区内斜坡地质灾害总数与研究区面积的比值。CF 接近 1 表示该因子作用下斜坡地质灾害发生的可能性较大,接近-1 表示该因子作用下斜坡地质灾害发生的可能性较小,接近 0 表示无法确定该因子与斜坡地质灾害发生与否二者之间的关系(Nocentini et al., 2024)。

2.4 随机森林模型 (RF) 介绍及构建

随机森林是分类和回归的集成学习算法,本研究主要运用其分类功能。该模型最早由 Breiman (Ksantini et al., 2024) 提出,其主要思路是利用 Bootstrap 自助重采样技术,从数据集集中有

放回地随机抽取样本,并分别建立相同数量 (k 个) 的决策树模型。每棵决策树根据数据集中纯度最高的特征作为划分依据,且每棵决策树单独工作,获得 k 种分类结果,再简单表决预测最终分类结果(Li L P and Lan H X, 2024)。

本文研究模型具体构建主要包括以下 4 步:

①选取等量的非灾害点,即基于 GIS 平台利用随机生成点功能生成 99 个非灾害点;②对评价因子进行相关性分析,同时通过各评价因子的信息量和系数值计算各因子对灾害发生的贡献程度;③将灾害点和非灾害点输入随机森林模型,得到易发性评价分区图;④统计各模型易发性分区面积及灾害点密度,同时绘制 ROC 曲线,分别计算其 AUC 值进行评估和对比,流程图如图 2 所示。

3 易发性评价

3.1 易发性评价指标体系的建立

基于 GIS 平台,选取全区 99 个斜坡地质灾害点作为样本数据,为了更合理地研究其发生的可能性,本次研究选择以 $30\text{ m} \times 30\text{ m}$ 的栅格为评价单元,全区栅格划分 4450 行,5110 列,共计 2273 万个评价单元。

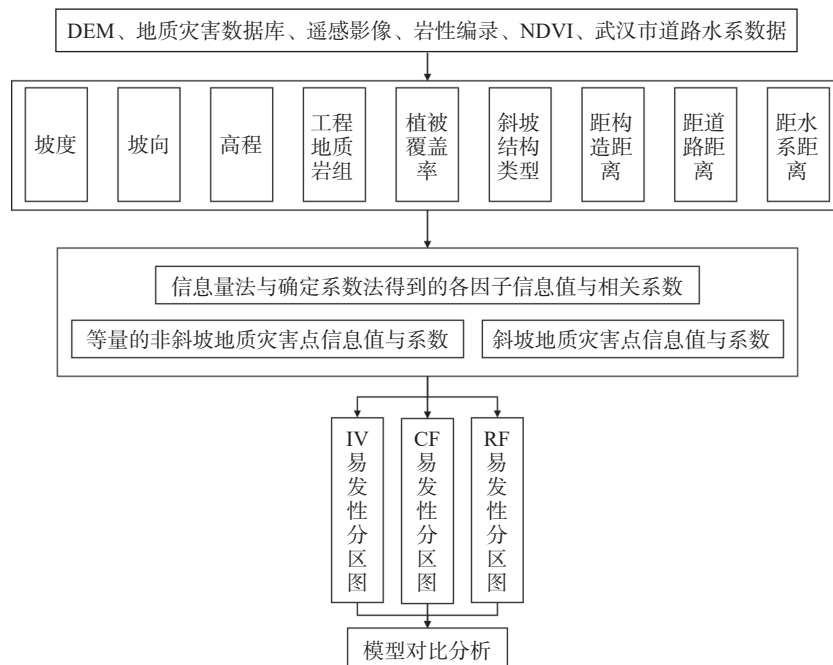


图 2 评价流程图

Fig. 2 Evaluation flow chart

据以上分析,研究区内斜坡地质灾害与地貌单元、地质构造、土层结构、人类工程活动等密不可分。为此选择坡度、坡向、高程、工程地质岩组、植被覆盖率、斜坡结构类型、距构造距离、距道路距离、距水系距离等 9 个评价指标建立研究区崩塌滑坡地质灾害易发性评价指标体系。

3.2 易发性指标因子状态划分

指标因子状态划分对后期评价模型的贡献率分析极其重要。本文选取的 9 个评价指标可分为两大类型:一是连续型指标,即该指标可通过 GIS 平台计算各个灾害点对应的唯一值,其中包括坡度、高程、距构造距离、距道路距离、距水系距离;二是离散型指标,其划分标准主要通过大量野外考察工作和前人的经验所决定。两种类型的指标,前者相对于后者来说,划分的区间需要结合实际情况具体分析。针对连续型指标,本文首先绘制灾害点在其上的频数分布直方图,依据灾害点的主要分布区间与对应区间所占的栅格数量两者的突变信息量确定划分区间的临界值。突变的信息量体现了研究区灾害点分布的特异性,能够提高后期评价模型的精度。

(1) 坡度

坡度是地质灾害发生的重要影响因素之一,坡度越大,岩石、土壤中产生的应力集中越严重,容易导致地质体的破裂和垮塌,同时地表水下渗和汇流速度越快,对岩石、土壤的侵蚀和冲刷作用越强,从而导致地质灾害的发生。基于 GIS 中的空间分析和多值提取到点功能,通过 DEM 数据提取坡度信息,以 5° 为一个间隔,分别统计灾害点的分布情况,其中坡度范围为 $0^\circ \sim 15^\circ$,灾害点分布较为集中,大于 35° 则无灾害点分布。因此,坡度分别划定四个区间为: $<5^\circ$ 、 $[5^\circ, 10^\circ)$ 、 $[10^\circ, 18^\circ)$ 、 $\geq 18^\circ$,同时计算灾害点密度(图 3a、图 4a、表 2)。

(2) 坡向

坡向是指坡面相对于正北方向的方位角,其对地质灾害的影响主要体现在光照:向阳坡受阳光照射多,岩石风化强烈,容易发生崩塌。背阳坡受阳光照射少,岩石风化较弱,但容易发生滑坡、泥石流等重力流灾害。运用 ArcGIS 中的表面分析功能,通过 DEM 数据提取坡向信息,分为平

面、北、东北、东、东南、南、西南、西和西北。统计灾害点在坡向上的分布情况,得出研究区斜坡地质灾害点主要集中东北、西南、南和北方向,共计 62 处,占灾害点总数的 61%(图 3b、图 4b、表 2)。

(3) 高程

高程可以控制坡体内应力值的大小,随着坡高的增加,应力值会显著增加。研究区总的地形为北高南低,以丘陵和平原相间的波状起伏地形为主,黄陂区北部和新洲区以东部分地区则有低山地形显示。基于灾害点的分布,将研究区划分为以下五个高程区间: $<25\text{ m}$ 、 $[25\text{ m}, 50\text{ m})$ 、 $[50, 100\text{ m})$ 、 $[100\text{ m}, 200\text{ m})$ 、 $\geq 200\text{ m}$ (图 3c、图 4c、表 1)。

(4) 工程地质岩组

工程地质岩组是指具有相似工程地质特征的一组岩石,这些特征包括岩石类型、结构、构造、风化程度、力学性质等,其是进行工程地质调查和评价的基础。根据地层岩性和野外调查数据,将研究区工程岩组分为坚硬片麻混合岩、片状层状变质岩、松散松软沉积、较坚硬碳酸盐岩、坚硬石英砂岩和层状碎屑沉积岩,其中松散松软沉积土体分布面积最广,为 7260.18 km^2 ,占研究区总面积的 84.7%;坚硬石英砂岩的灾害点密度最高为 $0.173\text{ 个}/\text{km}^2$ (图 3d、图 4d、表 2)。

(5) 植被覆盖率

植被覆盖率(NDVI, 归一化植被指数)是一种遥感指标,用于定量描述植被的覆盖程度和健康状况。作为常用评价指标之一,其变化可以反映地质环境的变化,植被覆盖率下降可能是地质灾害发生的前兆。在本研究,将归一化植被指数划分为 4 个区间,分别为: $<10\%$ 、 $[10\%, 30\%)$ 、 $[30\%, 60\%)$ 、 $\geq 60\%$ (图 3e、图 4e、表 1)。

(6) 斜坡结构类型

本研究按斜坡走向与岩层倾向的夹角将斜坡划分为以下六类:顺向坡:岩层倾向与坡向夹角 $<30^\circ$ 的斜坡;顺斜坡:岩层倾向与坡向交角在 $[30^\circ, 60^\circ)$ 之间的斜坡;横向坡:岩层倾向与坡向交角在 $[60^\circ, 120^\circ)$ 之间的斜坡;逆斜坡:岩层倾向与坡向交角在 $[120^\circ, 150^\circ)$ 之间的斜坡;逆向坡:岩层倾向与坡向交角在 $[150^\circ, 180^\circ)$ 之间的斜坡;平缓坡:包括块状结构的岩质边坡和土质边坡

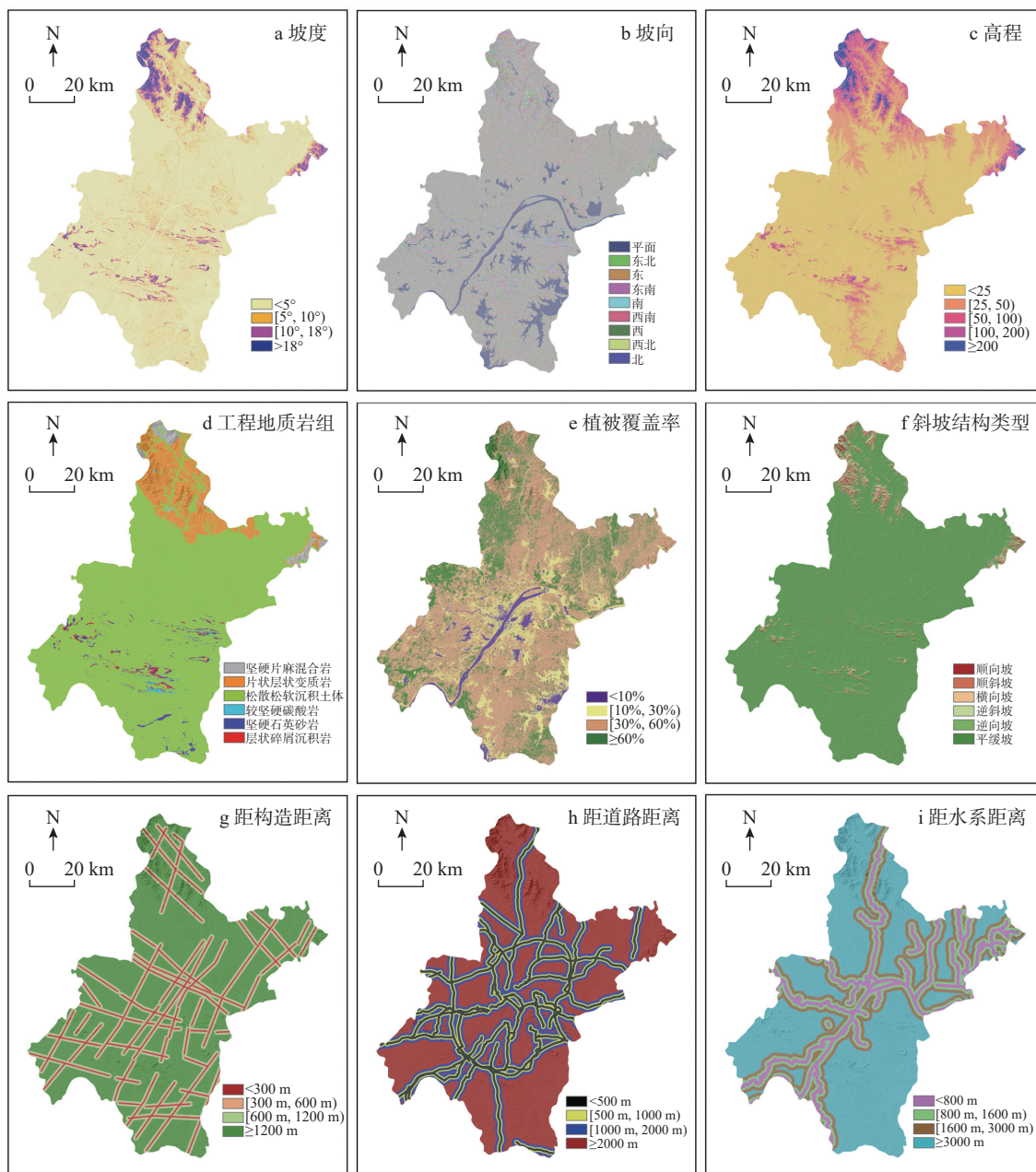


图 3 研究区评价因子分级图

Fig. 3 Grading chart of evaluation factors in the study area

(图 3f、图 4f、表 2)

(7) 距构造距离

断裂构造对于节理裂隙发育程度有重要影响。根据研究区的构造分布图,通过距构造距离构建缓冲区:分别为<300 m、[300 m, 600 m)、[600 m, 1200 m)、≥1200 m。统计斜坡地质灾害点在距离缓冲区的分布情况,得出在[300 m, 600 m)、[600 m, 1200 m)缓冲区间内灾害点分布最密集,灾害点密度分别为 1.172 个/km²和 0.617 个/km²(图 3g、图 4g、表 2)。

(8) 距道路距离

道路建设往往需要开挖边坡,这会破坏地质结构,降低边坡的稳定性。基于 GIS 平台,以主要道路(国道和省道)为中心构建缓冲区,划分 4 个缓冲区:<500 m、[500 m, 1000 m)、[1000 m, 2000 m)、≥2000 m。统计斜坡地质灾害点在距离缓冲区的分布情况,得出在<500 m、[500 m, 1000 m)缓冲区间内灾害点分布最密集,密度分别为 0.041 个/km²和 0.022 个/km²(图 3h、图 4h、表 2)。

(9) 距水系距离

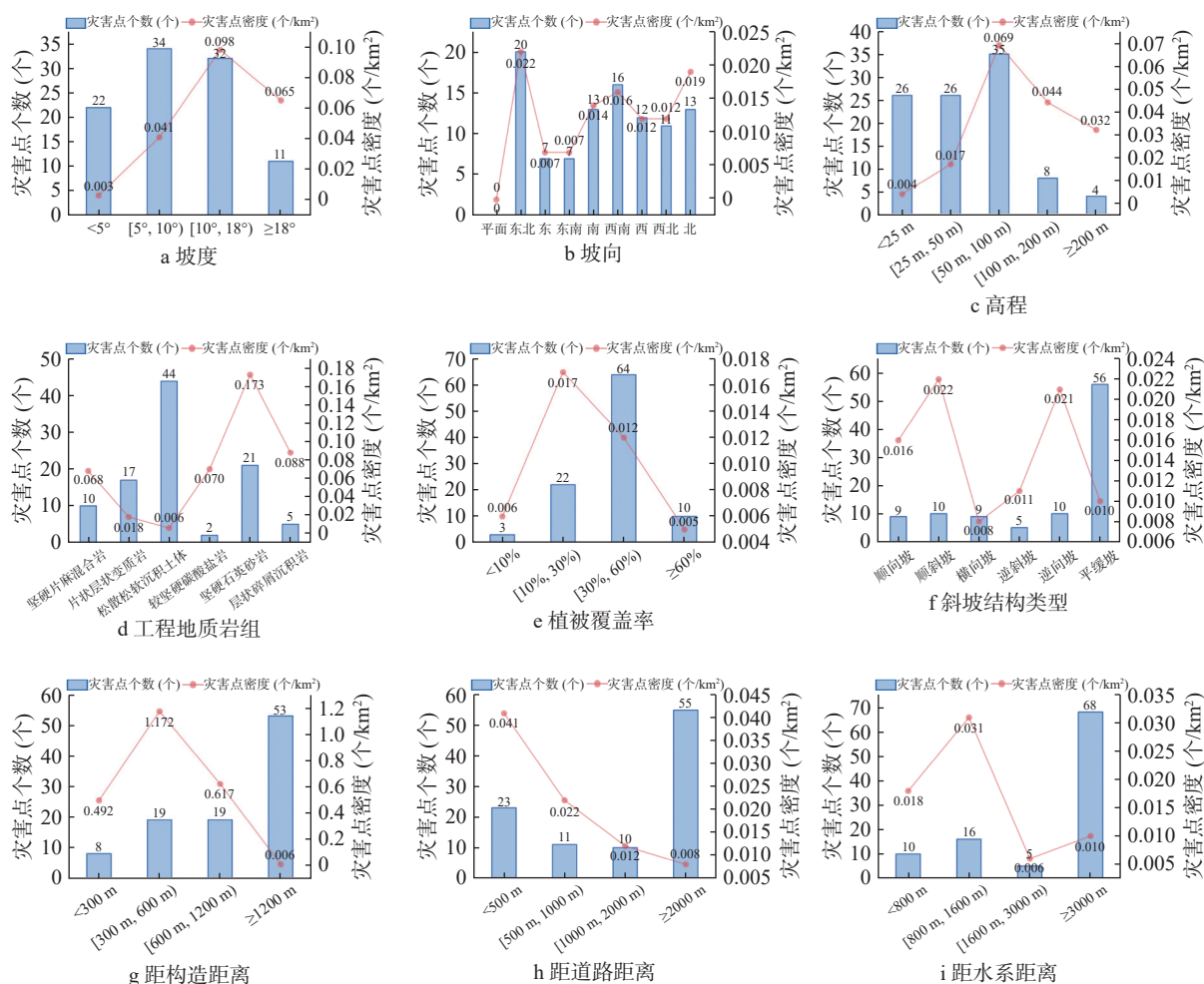


图4 研究区评价因子分级统计图

Fig. 4 Statistical chart of evaluation factor classification in the study area

基于 GIS 平台,以主要河流为中心构建缓冲区,划分 4 个区: <800 m、[800 m, 1600 m)、[1600 m, 3000 m)、≥3000 m。统计斜坡地质灾害点在距离缓冲区的分布情况,得出在 <800 m、[800 m, 1600 m) 缓冲区间上,灾害点分布最密集,密度分别为 0.018 个/km² 和 0.031 个/km² (图 3i、图 4i、表 2)。

4 易发性评价结果与验证

4.1 易发性评价结果

根据评价流程图(图 2),将表 2 中 9 个评价指标分级区间对应的信息量和系数分别赋值进行重分类,并将重分类后得到的栅格进行叠加分析得到研究区易发性指数区间,根据自然断点法,将研究区划分为 4 个易发性等级区间:低易发区、

中易发区、高易发区、极高易发区。对于 RF 模型,为了合理地选择非灾害点,以现有 99 处灾害点为中心构建 500 m 缓冲区,后使用 ArcGIS 的擦除功能,在研究区范围擦除缓冲区并与信息量模型评价结果中的低易发区进行叠加得到非灾害点选取的区域,在此区域上基于“创建渔网”工具生成 30 m × 30 m 间距的渔网点。创建好渔网点后,采用“子集要素”工具随机提取等量(即 99 处)的渔网点作为非灾害点。将非灾害点用 0 表示,灾害点用 1 表示,并使用随机抽样的方法对灾害点和非灾害点构建的样本集按 7 : 3 的比例分成训练集和测试集:其中训练集占 70%,测试集占 30%。之后采用网格搜索算法对模型进行参数优化和验证。如图 5a 所示,当模型的决策树数目达到 80 时,误差最低,之后随着决策树数目增加误差趋于稳定,因此设定决策树数目为 100 以提高

表 2 研究区评价因子分级统计表

Table 2 Statistical table of evaluation factor classification in the study area

评价因子	序号	因子分级	滑坡和 崩塌数(个)	占总数 比例(%)	面积 占比(%)	面积(km ²)	灾害点密度 (个/km ²)	信息量 IV	确定系数 CF
坡度	1	<5°	22	22.7	84.6	7253.74	0.003	-1.317	-0.722
	2	[5°, 10°)	34	35.1	9.8	837.62	0.041	1.277	0.737
	3	[10°, 18°)	32	33.0	3.6	307.50	0.098	2.187	0.901
	4	≥18°	11	11.3	2.0	170.28	0.065	1.742	0.839
坡向	1	平面	0	0.0	11.4	973.50	0.000	0.000	-1.000
	2	东北	20	20.2	10.6	910.76	0.022	0.642	0.479
	3	东	7	7.1	11.9	1020.33	0.007	-0.521	-0.409
	4	东南	7	7.1	12.0	1028.66	0.007	-0.529	-0.409
	5	南	13	13.1	11.2	956.93	0.014	0.162	0.151
	6	西南	16	16.2	11.6	998.08	0.016	0.328	0.283
	7	西	12	12.1	12.0	1030.02	0.012	0.008	0.008
	8	西北	11	11.1	11.1	949.95	0.012	0.002	0.002
	9	北	13	13.1	8.2	700.92	0.019	0.473	0.382
高程	1	<25 m	26	26.5	73.0	6253.01	0.004	-1.013	-0.619
	2	[25, 50 m)	26	26.5	17.7	1514.32	0.017	0.404	0.363
	3	[50, 100 m)	35	34.7	5.8	494.67	0.069	1.789	0.849
	4	[100 m, 200 m)	8	8.2	2.1	182.12	0.044	1.362	0.758
	5	≥200 m	4	4.1	1.5	125.03	0.032	1.006	0.663
工程地质岩组	1	坚硬片麻混合岩	10	10.1	1.7	148.13	0.068	1.765	0.846
	2	片状层状变质岩	17	17.2	11.1	953.82	0.018	0.435	0.387
	3	松散松软沉积土体	44	44.4	84.7	7260.18	0.006	-0.646	-0.446
	4	较坚硬碳酸盐岩	2	2.0	0.3	28.41	0.070	1.797	0.853
	5	坚硬石英砂岩	21	21.2	1.4	121.67	0.173	2.703	0.947
	6	层状碎屑沉积岩	5	5.1	0.7	56.93	0.088	2.038	0.884
植被覆盖率	1	<10	3	2.1	4.2	358.86	0.006	-0.693	-0.491
	2	[10%, 30%)	22	22.1	14.1	1206.20	0.017	0.449	0.372
	3	[30%, 60%)	64	66.3	61.3	5253.17	0.012	0.078	0.084
	4	≥60	10	9.5	20.4	1750.92	0.005	-0.764	-0.530
斜坡结构类型	1	顺向坡	9	9.2	6.8	580.35	0.016	0.302	0.294
	2	顺斜坡	10	10.2	5.3	455.20	0.022	0.655	0.505
	3	横向坡	9	9.2	13.0	1118.09	0.008	-0.346	-0.266
	4	逆斜坡	5	5.1	5.3	452.99	0.011	-0.038	0.003
	5	逆向坡	10	9.2	4.9	420.42	0.021	0.630	0.492
	6	平缓坡	56	57.1	64.7	5542.11	0.01	-0.125	-0.081
距构造距离	1	<300 m	8	8.1	0.2	16.27	0.492	3.701	0.989
	2	[300 m, 600 m)	19	19.2	0.2	16.21	1.172	4.564	1.000
	3	[600 m, 1200 m)	19	19.2	0.4	30.79	0.617	3.871	0.993
	4	≥1200 m	53	53.5	99.3	8505.89	0.006	-0.618	-0.431
距道路距离	1	<500 m	23	23.2	6.6	561.31	0.041	1.266	0.726
	2	[500 m, 1000 m)	11	11.1	5.7	491.94	0.022	0.660	0.489
	3	[1000 m, 2000 m)	10	10.1	9.9	847.04	0.012	0.022	0.022
	4	≥2000 m	55	55.6	77.8	6668.86	0.008	-0.337	-0.289
距水系距离	1	<800 m	10	10.1	6.5	555.68	0.018	0.443	0.362
	2	[800 m, 16000 m)	16	16.2	6.1	521.81	0.031	0.976	0.631
	3	[1600 m, 3000 m)	5	5.1	9.3	794.83	0.006	-0.608	-0.458
	4	≥3000 m	68	68.7	78.2	6696.83	0.010	-0.129	-0.122

模型精度。根据图 5b 所示, 9 个评价因子中坡度和高程与地质灾害的易发程度较为密切, 其特征重要性分别达到了 20.32% 和 20.11%, 坡向和距水系分布的特征重要性则相对较低, 分别为 2.76% 和 1.62%; RF 模型的统计指数如表 3 所示, 表中 TF 代表模型预测为灾害点, 实际为灾害点的样本个数; FP 代表模型预测为灾害点, 实际为非灾害点的样本个数; FN 代表模型预测为非灾害点, 实际为灾害点的样本个数; TN 代表模型预测为非灾害点, 实际为非灾害点的样本个数。对于准确率、

精确度、召回率和 F1 分数的值越高表明模型越好, 其中训练集和测试集的准确率分别为 0.993、0.900, 均大于 0.8, 其余指标也均接近 1。因此, 可以认为模型的预测结果具有较高的精度。

如图 6 所示, 三种模型均能正确反映研究区斜坡地质灾害的发育特征, 高易发区与极高易发区集中在研究区北部、中西部的高山地区及经河流剥蚀夷平形成的剥蚀堆积平原区域, 低易发区占绝大部分主要分布于长江沿岸和湖泊周边的平原地带及绝大部分低山丘陵地区。

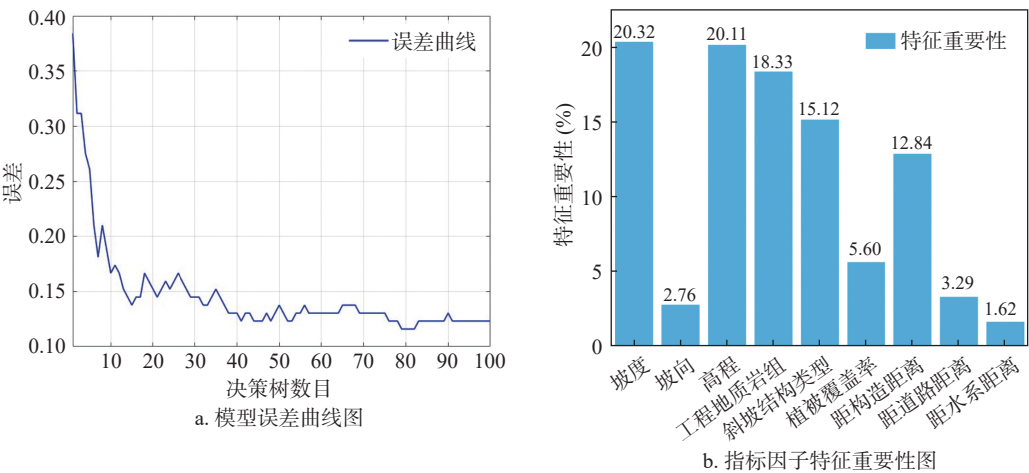


图 5 模型结果输出图
Fig. 5 Output plot of model results

表 3 RF 模型的统计指数表

统计指数	TF	FP	FN	TN	准确率	精确度	召回率	F1分数
训练集	64	1	0	73	0.993	0.985	1	0.99
测试集	32	3	3	22	0.900	0.914	0.914	0.91

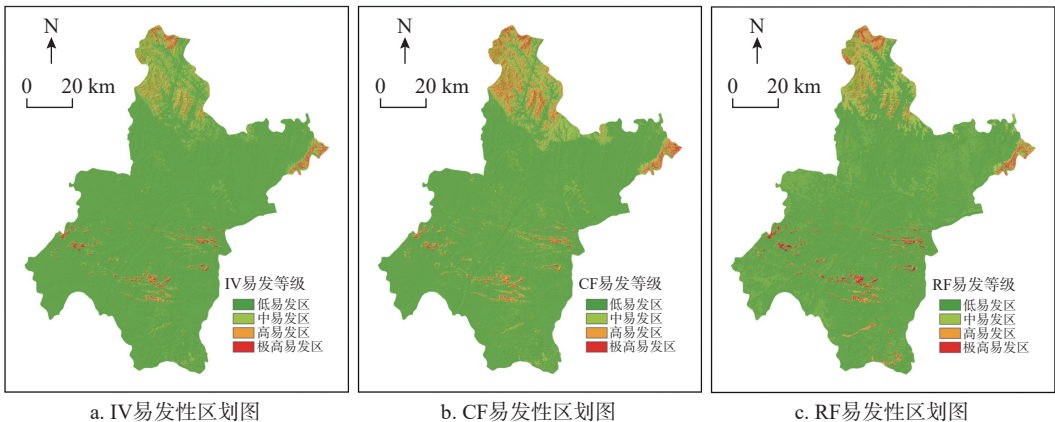


图 6 研究区地质灾害易发性区划图
Fig. 6 Geological hazard susceptibility zoning map of the study area

4.2 ROC 曲线精度验证

不同的模型使用相同的数据集进行分析会产生不同的分析结果,因此需采取科学方法对模型进行精度评估。采用受试者工作特征曲线(ROC)对区划结果进行精度检验能较好地衡量结果的可靠性,ROC 曲线越靠近左上角即曲线下方面积 AUC(Area Under Curve)值越接近 1,评价模型的精度越高。如图 7 所示,IV、CF、RF 模型对应的 AUC 值分别为 0.917、0.910、0.852,均大于 0.75。因此,三种模型均能正确反映研究区斜坡地质灾害的发育特征,同时模型精确度从高到低依次为 IV>CF>RF。

此外,为进一步比较三种模型区划结果之间的差异性,表 4 分别统计了各易发性分区面积及灾害点分布密度,其中 IV 和 CF 模型的极高易发区面积均小于 100 km²,灾害点密度均大于 0.2 个/km²,而 RF 模型的极高易发区面积则达到了 129.17 km²,

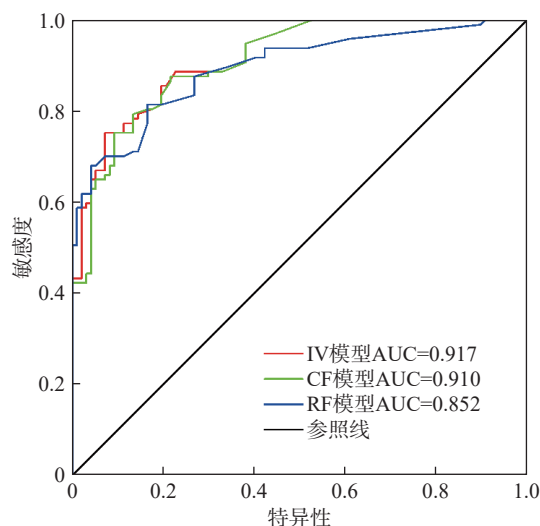


图 7 三种模型的 ROC 曲线图

Fig. 7 ROC plots for the three models

同时灾害点密度只有 0.170 个/km²。RF 模型的极高易发区在增大面积的情况下,灾害点密度仍小于另外两种模型,反映 IV 和 CF 模型的区划结果优于 RF 模型,与实际情况具有较高的贴合度。

表 4 研究区地质灾害易发性分区统计表

Table 4 Statistical table of geological hazard susceptibility in the study area

	灾害点个数			分区面积(km ²)			灾害点密度(个/km ²)		
	IV	CF	RF	IV	CF	RF	IV	CF	RF
低易发区	32	24	36	7612.492	7118.936	7216.09	0.004	0.003	0.005
中易发区	27	20	17	638.306	795.05	850.30	0.042	0.025	0.020
高易发区	19	33	24	224.471	562.309	373.58	0.085	0.059	0.064
极高易发区	21	22	22	93.881	92.854	129.17	0.224	0.237	0.170

5 结论

(1)三种模型均能正确反映研究区斜坡地质灾害的发育特征,高易发区与极高易发区集中在研究区北部、中西部的高山地区及经河流剥蚀夷平形成的剥蚀堆积丘陵区域;低易发区占绝大部分,主要分布于长江沿岸和湖泊周边的平原地带及绝大部分低山平原地区。

(2)IV 和 CF 模型二者得到的区划结果具有较高的相似度,同时二者的精确度及同等级易发区的灾害点密度均要高于 RF 模型。

(3)RF 模型在建模过程中分析了评价因子间的重要性,其中坡度、高程、工程地质岩组、斜坡

结构类型是权重较大的因子,而距水系距离、坡向、及距道路距离权重较低,说明研究区的斜坡地质灾害的发育因素主要与地形地貌、地层岩性有关。

参考文献:

- 陈建国,钟连祥.2023.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,34(6):105-115.
- 黄敏.2023.基于 FR 的多种机器学习模型在地质灾害易发性评价中的对比分析[J].福建地质,42(3):236-243.
- 焦伟之,张明,谢鑫鹏,李成文,刘涛,庞海松.2022.基于 GIS 与加权信息量模型的城镇地质灾害易发性评价——以大新镇为例[J].安全与环境工程,29(4):119-128.
- 连志鹏,厉一宁,刘磊,王宁涛.2022.基于 ArcGIS 的湖北省

- 远安县地质灾害风险定量评价[J]. 华南地质, 38(4):680-688.
- 刘坚,李树林,陈涛.2018.基于优化随机森林模型的滑坡易发性评价[J]. 武汉大学学报(信息科学版),43(7):1085-1091.
- 孟凡奇,高峰,林波,张京钊,徐京梁,牛伟.2023.基于AHP和信息量模型的地质灾害易发性评价——以鲁东片区为例[J]. 灾害学,38(3):111-117.
- 彭建兵,徐能雄,张永双,夏开文,薛翊国,张彬,杨国香,陈剑,王飞永,臧明东,梅钢,穆文平,王祚鹏,李星,张欣.2022.论地质安全研究的框架体系[J]. 工程地质学报,30(6):1798-1810.
- 邵大江,叶辉,王金亮,周京春,角媛梅,沙晋明.2023.基于机器学习均值化的地质灾害易发性评价[J]. 云南大学学报(自然科学版),45(3):653-665.
- 孙剑锋,马超,胡金树,闫铁生,臧加俊,徐辉.2023.基于灰色关联度与层次分析法耦合的地质灾害易发性评价——以浙江省云和县崇头镇为例[J]. 工程地质学报,31(2):538-551.
- 王雪冬,张超彪,王翠,朱永东,王海鹏.2022.基于Logistic回归与随机森林的和龙市地质灾害易发性评价[J]. 吉林大学学报(地球科学版),52(6):1957-1970.
- 温鑫,范宣梅,陈兰,刘世康.2022.基于信息量模型的地质灾害易发性评价:以川东南古蔺县为例[J]. 地质科技通报,41(2):290-299.
- 于喜坤,张紫昭,史光明,李崇博,刘毅业,朱建华,陈伟楠.2023.基于确定性系数与信息量耦合模型的新疆额敏县地质灾害易发性评价[J]. 工程地质学报, 31(4):1333-1349.
- 张威,胡舫瑞,蔡巍,彭琳,王咏林,陈枫.2023.基于XGBoost和云模型的地质灾害易发性评价[J]. 中国地质灾害与防治学报,34(6):136-145.
- 章昱,王磊,伏永朋,刘亚磊,黄皓.2023.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,39(3):512-522.
- 赵晓燕,李永平,谈树成.2022.GIS支持下CF与信息量耦合模型的攀枝花市矿山地质灾害易发性评价[J]. 云南大学学报(自然科学版),44(4):754-764.
- 翟文华,王小东,吴明堂,吴晓亮,李倩倩.2023.基于频率比模型和随机森林模型耦合的地质灾害易发性评价[J]. 自然灾害学报,32(6):74-82.
- Jiang N, Li Y E, Han Z, Yang J M, Fu B J, Li J Y, Li C L. 2024. A side-sampling based Linformer model for landslide susceptibility assessment: a case study of the railways in China[J]. Geomatics, Natural Hazards and Risk, 15(1): 2354507.
- Ksantini F, Sdiri A, Aydi A, Almeida-Naunay A F, Achour H, Tarquis A M. 2024. A comparative study based on AHP and fuzzy logic approaches for landslide susceptibility zoning using a GIS-based multi-criteria decision analysis[J]. Euro-Mediterranean Journal for Environmental Integration, 9(2): 1-20.
- Li L P, Lan H X. 2024. Analytical ‘decisiveness’ as a robust measure of the absolute importance of landslide predisposing factors[J]. International Journal of Digital Earth, 17(1): 2356161.
- Li R W, Tan S C, Zhang M F, Zhang S H, Wang H S, Zhu L. 2024. Geological Disaster Susceptibility Evaluation Using a Random Forest Empowerment Information Quantity Model[J]. Sustainability, 16(2): 765.
- Lombardo L, Opitz T, Ardizzone F, Guzzetti F, Huser R. 2020. Space-time landslide predictive modelling[J]. Earth-Science Reviews, 209: 103318.
- Nocentini N, Rosi A, Piciullo L, Liu Z Q, Segoni S, Fanti R. 2024. Regional-scale spatiotemporal landslide probability assessment through machine learning and potential applications for operational warning systems: a case study in Kvam (Norway)[J]. Landslides, 21(10): 1-19.
- Pradhan B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS[J]. Computers & Geosciences, 51: 350-365.
- Reichenbach P, Rossi M, Malamud B D, Mihir M, Guzzetti F. 2018. A review of statistically-based landslide susceptibility models[J]. Earth-Science Reviews, 180: 60-91.
- Sun K, Li Z Q, Wang S J, Hu R L. 2024. A support vector machine model of landslide susceptibility mapping based on hyperparameter optimization using the Bayesian algorithm: a case study of the highways in the southern Qinghai-Tibet Plateau[J]. Natural Hazards, 120(12): 11377-11398.
- Wang Z P, Chen J P, Lian Z C, Li F, Pang L, Xin Y B. 2024. Influence of buffer distance on environmental geological hazard susceptibility assessment[J]. Environmental Science and Pollution Research, 31(6): 9582-9595.
- Zhang S H, Tan S C, Wang H S, Shi Y Q, Ding D Y, Sun Y Q, Gao H X. 2024. To explore the optimal solution of different mapping units and classifiers and their application in the susceptibility evaluation of slope geological disasters[J]. Ecological Indicators, 163: 112073.