引文格式:郭飞,吴迪,葛民荣,等.连续变量因子分级和机器学习模型对滑坡易发性评价精度的影响[J].武汉大学学报(信息科学版),2026,51(2):236-248.DOI:10.13203/j.whugis20230413

Citation: GUO Fei, WU Di, GE Minrong, et al. Influence of Continuous Variable Factor Classification and Machine Learning Model on the Accuracy of Landslide Susceptibility Evaluation[J].Geomatics and Information Science of Wuhan University, 2026, 51(2):236-248.DOI:10.13203/j.whugis20230413

连续变量因子分级和机器学习模型对滑坡易发性评价精度的影响

郭 飞[1,2] 吴 迪[1,2] 葛民荣[3] 董进龙[4] 房 浩[5,6] 田东方[7]

1 三峡库区地质灾害教育部重点实验室,湖北 宜昌,443002
2 三峡大学土木与建筑学院,湖北 宜昌,443002
3 浙江省工程勘察设计院集团有限公司,浙江 宁波,315001
4 北京京岩工程有限公司,北京,100028
5 中国地质环境监测院,北京,100081
6 北京师范大学国家安全与应急管理学院,北京,100875
7 三峡大学水利与环境学院,湖北 宜昌,443002

摘 要:滑坡易发性评价建模中环境因子分级区间和机器学习模型对建模结果的影响不容忽视。为探究这两种因素对滑坡易发性评价结果的影响规律,基于主客观赋权法通过对环境因子进行赋权以构建评价指标体系,再利用地理探测器探究不同连续变量因子分级对滑坡易发性评价结果精度的影响规律,进而分别采用随机森林模型、梯度极限提升模型和遗传算法优化的神经网络模型开展滑坡易发性研究。结果表明:通过地理探测器得到的与灾害关联度最高的分区组合计算出的最大曲线下面积值为0.886,说明该方法可以得到最优的分级区间,且能有效提高易发性评价结果的精度;在易发性评价结果中,随机森林模型最优,较梯度极限提升模型和遗传算法优化的神经网络模型精度分别提高了9.7%和9.6%。基于地理探测器的环境因子最优分级区间是合理的,且随机森林模型作为滑坡易发性评价模型是高效准确的。

关键词:滑坡易发性;地理探测器;主客观组合赋权;连续变量因子分级;机器学习模型

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-03-11

DOI:10.13203/j.whugis20230413 **文章编号**:1671-8860(2026)02-0236-13

Influence of Continuous Variable Factor Classification and Machine Learning Model on the Accuracy of Landslide Susceptibility Evaluation

GUO Fei[1,2] *WU Di*[1,2] *GE Minrong*[3] *DONG Jinlong*[4] *FANG Hao*[5,6] *TIAN Dongfang*[7]

1 Key Laboratory of Geological Hazards on Three Gorges Reservoir Area, Ministry of Education, Yichang 443002, China
2 College of Civil Engineering and Architecture, China Three Gorges University, Yichang 443002, China
3 Zhejiang Engineering Survey and Design Institute Group Co, Ltd. Ningbo 315001, China
4 Beijing Jingyan Geotechnical Co, Ltd. Beijing 100028, China
5 China Geological Environment Monitoring Institute, Beijing 100081, China
6 School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China
7 College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China

Abstract: **Objectives**: The influence of environmental factor classification interval and machine learning model on modeling results in landslide susceptibility evaluation modeling cannot be ignored. In order to explore the influence of these two factors on the evaluation results of landslide susceptibility. **Methods**: The evaluation index system is constructed by weighting the environmental factors based on the subjective and

基金项目:国家自然科学基金(42107489);国家重点研发计划(2021YFC3001901、2022YFC3005603);湖北巴东地质灾害国家野外科学观测研究站开放基金(BNORSG202304);三峡库区地质灾害教育部重点实验室开放基金(2022KDZ14);湖北省自然科学基金(2022CFB557);土木工程防灾减灾湖北省引智创新示范基地(2021EJD026)。

第一作者:郭飞,博士,副教授,主要从事区域地质灾害风险评估研究。ybbnui.2008@163.com

通信作者:田东方,博士,教授。tdf_2005@163.com

objective weighting method, and then the influence of different continuous variable factor classifications on the accuracy of landslide susceptibility evaluation results is explored by using the GeoDetector. Then, the random forest model, the gradient boosting model and the neural network model optimized by genetic algorithm are used to study the landslide susceptibility. **Results:** The results show that: the maximum area under curve value calculated by the partition combination with the highest correlation degree with the disaster obtained by the GeoDetector is 0.886.This finding indicates that the method can obtain the optimal classification interval and can effectively improve the accuracy of the susceptibility evaluation results. In the evaluation of susceptibility, the random forest model has been demonstrated to be the most effective, with a 9.7% and 9.6% increase over the gradient limit lifting model and the neural network model optimised by genetic algorithm, respectively. **Conclusions:** The optimal classification interval of environmental factors, as determined by GeoDetector, is reasonable, and the random forest model is both efficient and accurate in its evaluation of landslide susceptibility.

Key words: landslide susceptibility; GeoDetector; subjective and objective combination empowerment; continuous variable factor classification; machine learning model

滑坡作为一种频发的地质灾害,每年在世界范围内造成重大的人员伤亡和经济损失[1]。滑坡易发性评价作为防灾减灾的有效途径之一,一直是滑坡领域研究的热点问题[2]。滑坡易发性是指特定地点处斜坡演化为滑坡的空间概率,主要包括滑坡编目、环境因子、评价单元、评价模型、联接方法和精度评价等方面[3]。其中环境因子和评价模型是滑坡易发性评价中的重要环节,直接影响评价结果的精度。滑坡易发性评价作为滑坡灾害预测预防的基础,如何有效地提高滑坡易发性评价结果的精确性显得尤为重要[4]。

引起地质灾害的因素可分为内在因素和外部诱发因素,这些因素又可以细分为许多不同的因子。各个地区的地形条件、地质背景、气候条件等存在差异,因此区域内的环境因子组合可能也会不同。目前对于环境因子的构建并无标准或规范可依,大都通过前人的研究经验,总结出一系列原则来确定。当前的筛选方法包括层次分析法(analytic hierarchy process,AHP)[5]、主成分分析法(principal component analysis,PCA)[6]、递归特征消除法[7]、逐步回归法[8]、地理探测器[9-10]以及权重均值法[11]等。由于环境因子作为易发性研究的输入数据,直接影响易发性评价结果的精确度。因此,针对特定区域选取合适类型和数量的因子对滑坡易发性评价结果的可靠性起至关重要的作用。

研究表明,同一环境因子不同属性区间,对易发性结果的影响是不同的,因而有部分学者对滑坡易发性评价出现的连续变量因子和非连续变量因子分别进行分级和分类研究,试图探寻其规律所在。文献[12]将12个连续变量的因子分级为4、8、12、16和20,计算易发性之后,发现分级从4增加至8,易发性预测结果的精度快速增加,而分级从8增加到20时,易发性预测精度增速放缓直至趋于稳定;文献[13]提出了对称分类法对非连续变量因子进行分类,并应用多种分类方法进行滑坡易发性评价,结果表明对称分类法得出的效果最佳;文献[14]分别对13个非连续变量使用自然断点法进行因子分类,并通过两种分类结果构建模型预测研究区内滑坡易发性,发现采用滑坡点属性确定因子分类标准建立的随机森林(random forest,RF)模型精确性和预测能力更优。以上研究表明,环境因子的区间划分亦对滑坡易发性评价结果产生较大的影响[15]。

目前滑坡易发性评价模型较多,主流的方法均基于统计分析,主要有经典统计方法和基于机器学习方法[16]。考虑到滑坡发生机理的复杂性,经典统计方法往往不够准确,因而现多选用精度更高的机型学习模型方法[17-18]。而关于模型的选用尚无统一标准,不同的模型具有各自的优点和不足[19]。机器学习模型作为一种数据驱动模型,凭借其对大数据优越的处理能力在易发性建模领域被广泛应用[20]。文献[21-22]认为反向传播神经网络(back propagation neural network,BPNN)模型能掌握滑坡的内在变化规律,计算思路较为科学合理,对滑坡易发性结果预测更佳;文献[23-24]发现,采用梯度极限提升模型(extreme gradient boosting,XGBoost)获得的滑坡易发性结果预测精度高、稳定性好;更有研究表明[25-29],RF模型因其具有无需考虑数据尺度和数据分布、计算效率高、评价结果精度高且超参数调试成本较神经网络、深度学习模型明显较低、

泛化能力强等优势被广泛采用[30]。综上所述,机器学习模型能有效解决统计模型预测区域滑坡易发性时存在的问题,而BPNN、XGBoost和RF模型优势明显,从而选择这3类模型开展对比研究。

针对当前在滑坡易发性评价中连续变量因子分级研究的不足以及机器模型结果的较大差异,本文通过对主客观组合赋权法[31]得到连续变量因子,在进行不同分级后,结合现有研究中优势明显的BPNN、XGBoost和RF模型探究其所得到的易发性评价结果的影响规律,并给出连续变量因子的最优分级和所选取的易发性评价模型中预测精度最高的模型,再结合宁都县赖村镇验证研究结果的可靠性。研究结果可为滑坡的易发性研究提供一种新思路,并为当地滑坡防治工作提供参考依据。

1 研究方法

1.1 研究思路

本文通过研究环境因子的筛选,进而探究连续变量因子最优分级区间以及不同机器学习模型对滑坡易发性评价结果的影响规律,其具体研究思路如图1所示,包括以下几个方面:

1)基于历史数据和现场调查,收集研究区的地形地貌、地质环境、气象水文和人类工程活动等背景资料;

2)基于ArcGIS提取基础环境因子,通过主客观组合赋权法对基础环境因子进行排序,从而根据权重筛选环境因子以得到关联性最高的连续变量因子;

3)利用地理探测器(GeoDetector,GD)[32]计算不同分级区间的连续变量因子与灾害点的关联度,通过模型计算检验不同分级区间对易发性结果的影响,并找出最优的环境因子分级区间;

4)通过在不同的机器学习模型中计算易发性结果,根据精度统计分析和引入受试者工作特征曲线(receiver operating characteristic curve,ROC)以及曲线下面积(area under curve,AUC)检验易发性结果,比较BPNN[33]、XGBoost[34]、RF[35]3种不同机器学习模型的精度,确认精度最高模型。

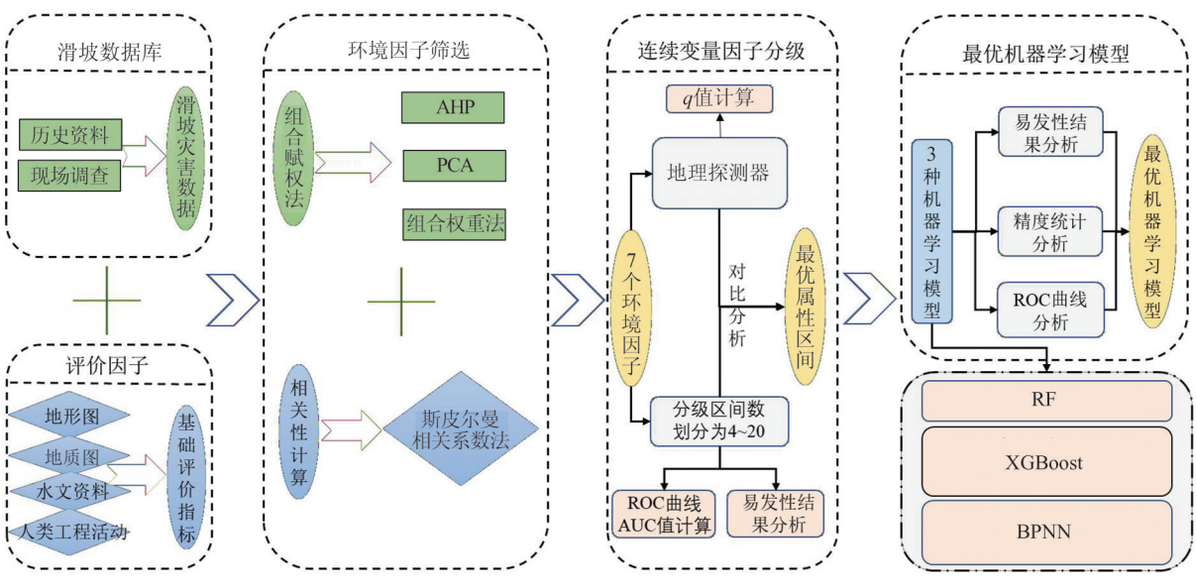


图1 模型构建流程图

Fig. 1 Flowchart of Model Construction

1.2 GD原理

GD是一种空间统计工具,该工具利用滑坡的空间分层异性和评价因子对滑坡的空间分层异性的解释程度,可判定潜在因素对滑坡产生的贡献大小,度量指标为因子驱动力q值,计算式为:

$$q=1-\frac{1}{n\sigma^2}\sum_{h=1}^{m}n_h\sigma_h^2=1-\frac{E_1}{E_2} \tag{1}$$

$$E_1=\sum_{h=1}^{m}n_h\sigma_h^2 \tag{2}$$

$$E_2=\frac{1}{n\sigma^2} \tag{3}$$

式中,σ^2与n为h的栅格值方差和栅格总数;σ_h^2、n_h和m分别为第h类栅格值方差、栅格数和类型总数;E_1为全区的方差;E_2为层方差之和。q值范围为0~1,q=0表示潜在诱发因子与滑坡的空间分

布完全不具有相似性,$q=1$表示潜在诱发因子与滑坡的空间分布一致。根据该探测值可确定环境因子不同分级区间与滑坡灾害的空间相似程度。

1.3　滑坡易发性评价模型

1)BPNN模型

BPNN模型是一种按误差逆向传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一,由信息的正向传播和误差的反向传播两个过程组成[36]。BPNN模型大约占了神经网络算法使用的75%以上,具有强大的非线性能力和映射能力。其模型计算式为:

$$z_j=f\left(\sum_{i=1}^{m}w_{ij}x_i+b_j\right) \tag{4}$$

$$y_k=f\left(\sum_{j=1}^{n}w_{jk}z_j+b_k\right) \tag{5}$$

式中,x_i为第i个输入层;z_j为第j个隐含层;y_k为第k个输出层;w_{ij}、w_{jk}分别为输入层与隐含层、隐含层与输出层之间权值;f为激励函数;b_j、b_k分别为节点阈值。

2)XGBoost模型

XGBoost模型是一种新型的梯度提升决策树算法,同样是串行地生成模型,取所有模型的和为输出。XGBoost模型将损失函数作二阶泰勒展开,利用损失函数的二阶导数信息优化损失函数,根据其是否减小来选择是否为分裂节点[20]。在处理数据缺失值方面也做了一定的优化。其模型计算式为:

$$\theta(x)=\sum_{j=1}^{T}\left[\left(\sum g_i\right)w_j+\frac{1}{2}\left(\sum h_i+\lambda\right)w_j^2\right]+\gamma T \tag{6}$$

式中,θ为目标函数;γ、λ为正则化系数;g_i为损失函数l的一阶偏导数;h_i为损失函数l的二阶偏导数;T为叶子数;w_j为叶子节点值。

3)RF模型

RF模型是通过不同的数据子集构建多个决策树,然后对多个决策树的判断结果进行投票,最终得到输出结果。RF算法每次都是通过随机抽取训练样本集的一部分和随机选择被抽取样本的一部分特征进行单个学习器训练,这样既能保证每棵树相互独立,又使投票结果更准确[37]。其模型计算式为:

$$\hat{f}(x)=\frac{1}{k}\sum_{i=1}^{k}t_i(x) \tag{7}$$

式中,$\hat{f}(x)$为回归树预测结果;k为回归树数量;$t_i(x)$为第i个回归模型计算结果。

2　研究区与数据

2.1　研究区概况

赖村镇位于江西省赣州市宁都县西南部,总面积为180 km^2。赖村镇地貌以中低山、丘陵为主,西北部地势较高,总体为中低外高,海拔在800 m左右。研究区地处亚热带湿润季风气候,年平均降水量为1 583.9 mm,年平均气温在19.2 ℃。全镇共有河流20多条,其中青塘河自北向南纵贯赖村镇,在于都县流入贡江,境内流长25 km,为境内主干河。此外研究区地层出露比较齐全,主要有前寒武纪变质岩、泥盆纪-侏罗纪沉积盖层、白垩纪红色碎屑岩和第四系松散沉积层。宁都县赖村镇矿产资源丰富,由于矿山开采形成大量的人工切坡,同时村民因建房亦形成大量人工切坡,在降雨等因素作用下,严重威胁当地村民住房和出行安全,因此开展赖村镇滑坡的易发性评价可为当地国土空间规划、滑坡防治工作提供参考,其滑坡灾害的空间位置分布如图2所示。

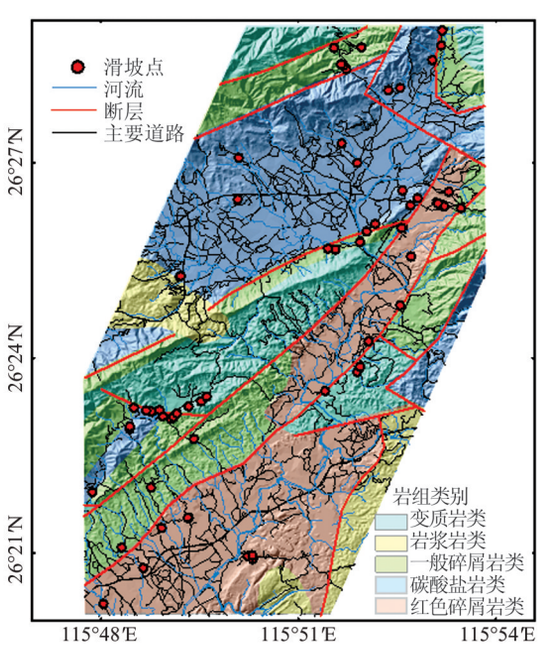


图2　研究区概况

Fig. 2　Overview of the Study Area

该地滑坡规模均为小于1×10^5 m^{3}的小型新滑坡,其中大部分为土质(86.7%)滑坡,少量为碎块石(8.3%)和岩质滑坡(5%)。同时大多数为平移式浅中层滑坡(<25 m),为人类工程活动所致。

2.2　数据来源

本研究通过现场调查的方式建立部分滑坡

灾害数据库。其中基础评价指标及与滑坡相关的数据主要来源有以下几个方面:①1∶10 000地形图,主要用于数字高程模型(digital elevation model,DEM)的生成和地形地貌、水文特征等连续性指标的提取;②1∶10 000地质图,主要用于地层岩组、构造断层和坡体结构等地质指标的提取;③基于滑坡灾害及隐患调查表建立的滑坡编录,主要用于滑坡发育分布特征分析以及作为样本数据和各评价指标建立联系。

3 易发性评价指标体系构建

根据现场调查分析,工作区滑坡灾害主要受地形地貌、地质环境、水文、人类工程活动、植被覆盖等因素的影响,因此本研究基于ArcGIS平台和遥感等技术方法共提取了15个环境因子展开研究。

3.1 主客观组合赋权

目前环境因子的筛选主要依据研究者的经验或者数学工具计算权重来进行,前者会存在较强的主观性,忽略影响灾害的次要因素,后者的计算对数据的要求很大,数据不足或数据精度不够都会产生计算错误。针对这些问题,本文通过主客观组合赋权的方法来选取环境因子。

通过PCA[38]和AHP[39]对提取的15个环境因子进行计算,得到最终各指标权重结果如图3所示。由图3可以看出,PCA求得的权重分布较为平均,地形地貌因子(坡度、粗糙度、地面起伏度、坡向)和地质条件因子(地层岩性、断层距离)权重较大,而与滑坡灾害关系最为密切的人类活动因子(道路距离、房屋密度)的权重并未排在前列,说明客观方法倾向于认为地形地貌因子和地质条件因子对研究区的滑坡灾害影响较大。而AHP求得的权重较为集中,道路距离、房屋密度和地层岩性占总权重的65.2%,在空间上主要分布在距离道路沿线和房屋周围,滑坡灾害受人类工程建设活动影响较大。除了道路距离、房屋密度和地层岩性3个评价因子之外,其他的评价因子的权重较低,反映了主观方法对控制滑坡灾害主要因素的侧重,而弱化了其他因素的影响。

主客观组合赋权[40]是将主观、客观权重中的各种环境因子的权重通过一定方法求出它们在综合权重中的占比,而后按照占比加权取和为最终的综合权重,其计算式为:

$$W=\mu W_1+\gamma W_2 \tag{8}$$

式中,W_1为通过AHP计算得到的主观权重;W_2为PCA得到的客观权重;μ为主观权重的加权系数;γ为客观权重的加权系数。通过距离函数分配主客观权重占比的基本思想是使不同权重的差异程度和其分配系数之间的差异程度达到一致,具体方法如下:设W_1、W_2的距离函数为$d(W_1,W_2)$,则:

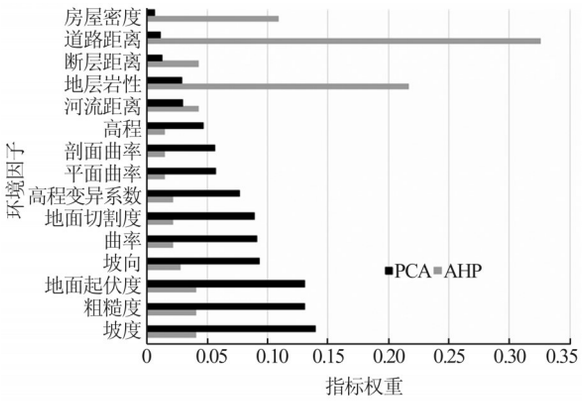


图3 PCA与AHP分析排序图

Fig. 3 PCA and AHP Analysis Ranking Diagram

$$d(W_1,W_2)=\left[\frac{1}{2}\sum_{i=1}^{n}(W_1-W_2)^2\right]^{\frac{1}{2}} \tag{9}$$

$$d(W_1,W_2)^2=(\mu-\gamma)^2 \tag{10}$$

$$\mu+\gamma=1,\mu\geqslant 0,\gamma\geqslant 0 \tag{11}$$

组合权重结果如表1所示,其中μ=0.615,γ=0.385。可以看出,基于主客观组合赋权的方法不仅突出了研究区内影响滑坡灾害的主要因素:人类工程建设活动因素(道路距离、房屋)和地质条件因素(地层岩性、断层距离),还给地形地貌因素分配了一定的权重,并突出了地形地貌类因子中比较重要的坡度、粗糙度等因子。

表1 组合权重结果

Table 1 Combination Weight Results

环境因子	W_1	W_2	W
道路距离	0.326	0.077	0.230
地层岩性	0.217	0.131	0.184
房屋密度	0.109	0.047	0.085
坡度	0.041	0.140	0.079
粗糙度	0.041	0.131	0.076
断层距离	0.043	0.093	0.062
地面起伏度	0.041	0.091	0.060
坡向	0.028	0.089	0.051
河流距离	0.043	0.056	0.048
地面切割度	0.022	0.057	0.035
平面曲率	0.015	0.029	0.021
剖面曲率	0.015	0.029	0.020
高程变异系数	0.022	0.012	0.018
曲率	0.022	0.011	0.018
高程	0.015	0.006	0.012

相较于主观方法,组合赋权的方法还合理地分配了权重给次要因素;相较于客观方法,组合赋权方法排序的结果与研究区灾害发育的实际情况更为贴切。

3.2　指标筛选

在探讨环境因子最优分级区间时需选用连续型环境因子,而地面切割度、地面起伏度、高程、高程变异系数、粗糙度、坡度、坡向、平面曲率、剖面曲率和曲率10个环境因子均通过DEM提取,具有一定的关联性。因此,采用斯皮尔曼相关系数法[41]分析这10个环境因子之间的相关性,其计算结果如表2所示。

由于高程、地面起伏度、高程变异系数和地面切割度都是基于高程换算求得,具有较强的相关性,所以只能将其中一个划入指标体系中。由表2结果可知,高程与地面起伏度、高程变异系数和地面切割度3个因子之间的相关系数较高,分别为0.452、0.962、0.759;高程变异系数与地面起伏度、地面切割度之间的相关系数分别为0.433、0.902;地面起伏度和地面切割度的斯皮尔曼相关系数为0.476。综合来看,地面起伏度和其他3个因子的相关系数最小,因此保留地面起伏度因子。而曲率、平面曲率和剖面曲率3个因子之间的相关系数很高,因此不采用。

综上所述,结合主客观组合赋权法选取的权重最大的8个环境因子以及斯皮尔曼相关系数法选取的相关性较低的7个环境因子,除去地层岩性离散型环境因子外,选取了道路距离、房屋密度、断层距离、坡度、粗糙度、坡向和地面起伏度7个连续变量因子开展其最优分级研究。

表 2　10个DEM指标的斯皮尔曼相关系数

Table 2　Spearman Correlation Coefficients of Ten DEM Indices

DEM指标	地面切割度	地面起伏度	高程	高程变异系数	粗糙度	坡度	坡向	平面曲率	剖面曲率	曲率
地面切割度	1.000	0.476	0.759	0.902	0.891	0.292	0.453	0.923	0.646	0.879
地面起伏度		1.000	0.452	0.433	0.504	0.305	0.245	0.535	0.370	0.439
高程			1.000	0.962	0.402	0.203	0.296	0.611	0.474	0.781
高程变异系数				1.000	0.344	0.445	0.457	0.454	0.690	0.626
粗糙度					1.000	0.276	0.197	0.583	0.423	0.837
坡度						1.000	0.101	0.588	0.225	0.637
坡向							1.000	0.495	0.525	0.562
平面曲率								1.000	0.975	0.839
剖面曲率									1.000	0.816
曲率										1.000

4　连续变量因子最优分级区间探究

4.1　因子分级区间驱动力计算

基于GD原理可知,驱动力q值越大,表明该因子与灾害的关联性越大,通过计算分析7个连续变量因子如图4所示,得到不同因子分级区间的q值如表3所示,因子最大q值对应的最优分级区间如图5所示。由以上结果可以看出,不同因子的最优分级区间并不相同。除房屋密度外,其余6个因子的驱动力q值都是随着分级区间的增加呈现出先增长到最大值后减弱的趋势,由于本次选取的因子分级区间为4~20,选取的最大分级区间有限,从而导致房屋密度环境因子的q值最大在分级区间为20时呈现,因此综合分析认为,当q值最大时,此时环境因子处于最优分级区间,与灾害点的关联度最高。

4.2　最优分级区间验证

从易发性计算结果的角度,由于RF模型被普遍认为是目前预测精度较高的模型,因此本研究直接利用RF模型计算在各个分级区间下的易发性。再通过GD得到的与灾害关联度最高的分级组合设为第18种易发性分区结果,然后将这18种易发性结果。按照中国地质调查局发布的《地质灾害风险调查评价规范(1:50 000)》(DZ/T 0438－2023)[42]标准要求,将区域滑坡易发程度划分为4个等级,通过ROC检验其是否为最优。计算得到的结果如图6所示,其中r为环境因子分级区间。

通过计算得到18种易发性结果,得到这18种情况下的易发性结果ROC曲线如图7所示。由图7可以看出,通过GD得到的与灾害关联度最高的分区组合计算出的最大AUC值为0.886,说明该方法可以得到最优的分级区间,且能有效提高易发性评价结果的精度。

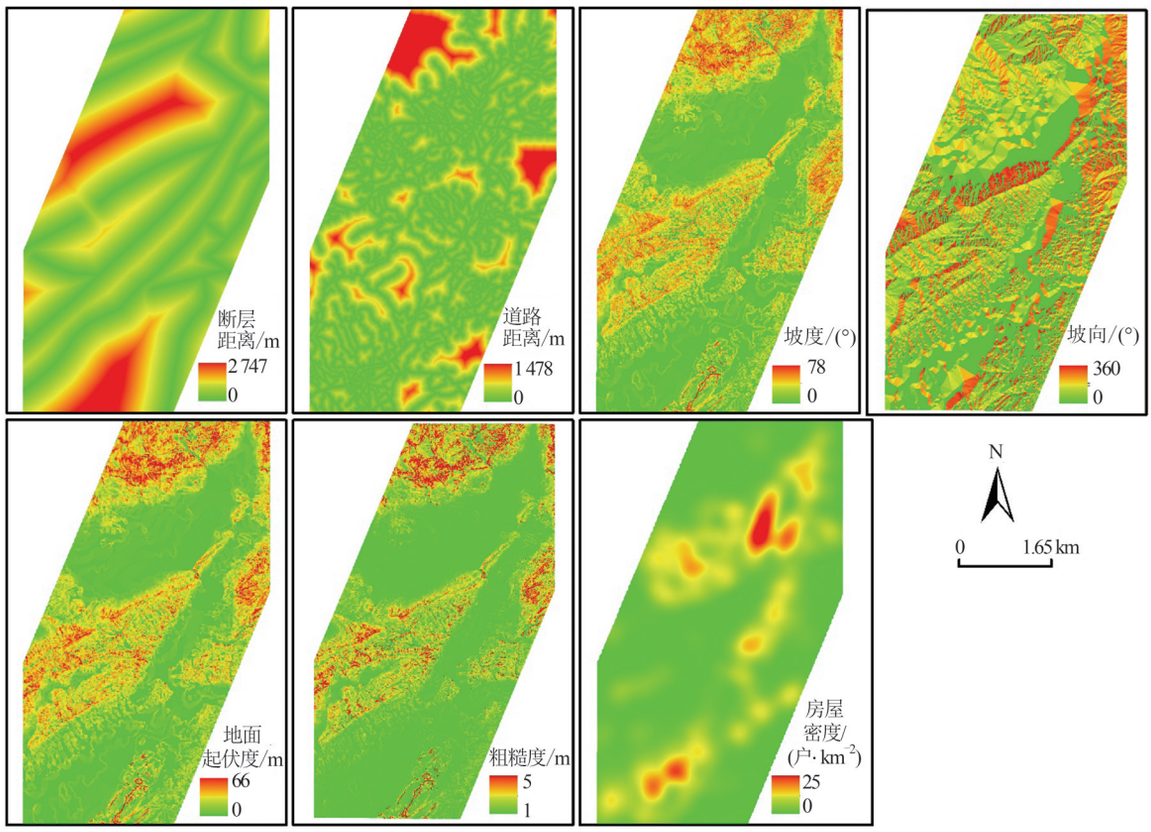


图4 7个连续型环境因子结果

Fig. 4 Results of Seven Continuous Environmental Factors

表3 环境因子分级区间q值

Table 3 Environmental Factor Grading Interval q Value

属性区间数	坡度	坡向	粗糙度	地面起伏度	房屋密度	道路距离	断层距离
4	0.036	0.032	0.051	0.023	0.013	0.048	0.030
5	0.057	0.043	0.056	0.050	0.041	0.049	0.041
6	0.061	0.045	0.069	0.060	0.056	0.059	0.053
7	0.040	0.044	0.070	0.056	0.065	0.055	0.055
8	0.054	0.041	0.072	0.054	0.067	0.059	0.062
9	0.061	0.062	0.057	0.054	0.040	0.056	0.058
10	0.061	0.062	0.057	0.055	0.043	0.059	0.059
11	0.062	0.065	0.057	0.059	0.048	0.059	0.061
12	0.058	0.067	0.057	0.060	0.048	0.058	0.064
13	0.065	0.069	0.057	0.055	0.056	0.058	0.062
14	0.068	0.072	0.057	0.055	0.062	0.065	0.062
15	0.067	0.069	0.057	0.071	0.062	0.067	0.064
16	0.065	0.069	0.057	0.072	0.075	0.063	0.066
17	0.065	0.066	0.057	0.071	0.075	0.061	0.067
18	0.061	0.062	0.057	0.071	0.075	0.059	0.066
19	0.056	0.065	0.057	0.070	0.085	0.060	0.066
20	0.064	0.066	0.057	0.065	0.088	0.063	0.065

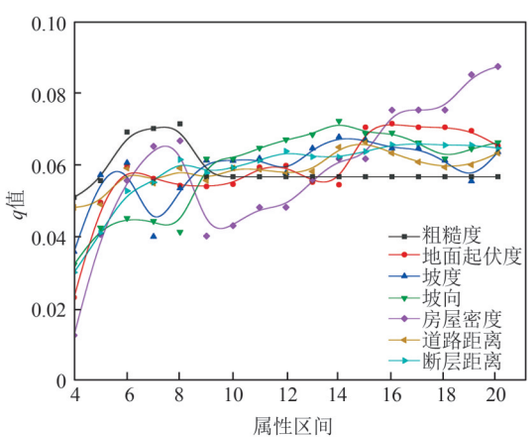


图5 环境因子最大q值对应最优分级区间分布

Fig. 5 The Maximum q Value of Environmental Factor Correspondsing to the Optimal Grading Interval Distribution

4.3 对比分析

与文献[12]的研究不同的是,此次设置了连续整数区间数,讨论了更多种分级区间的情况。不同分级区间数与易发性结果的关系如图8所示,其中黑点为不同分级区间数对应的AUC值,红色的三角符号是GD算出的分级区间,可以看出GD计算的分级区间比其他各种分级区间的AUC值都要高得多。

文献[12]的结果显示,分级区间数8是一个分界点,分级区间数小于8时,随着分级区间数的增加,AUC值增加得较明显,分级区间数大于8时,随着分级区间数的增加,AUC值增加的较慢,

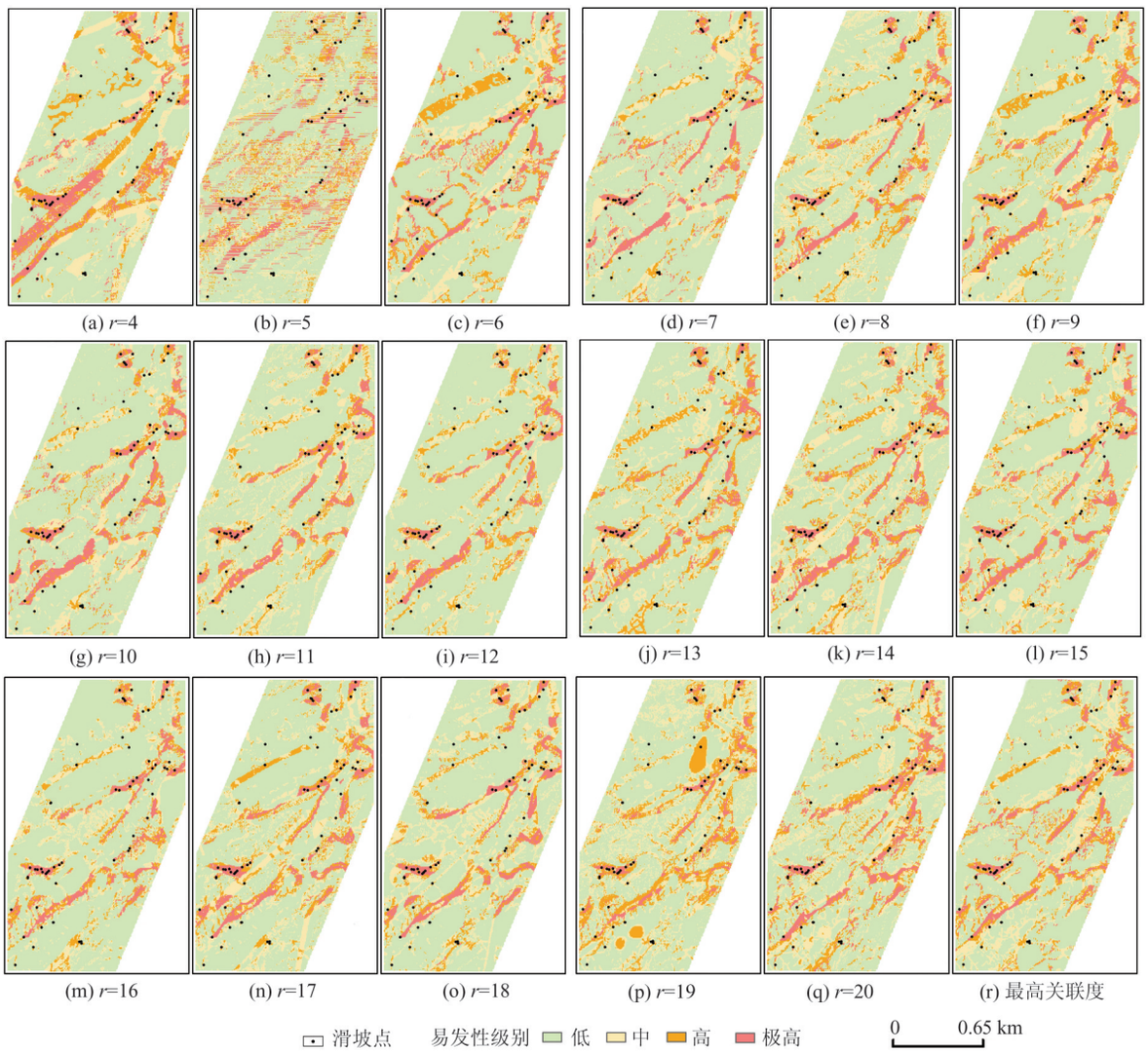


图 6　环境因子不同分级区间 r 和最高关联度的易发性结果

Fig. 6　Susceptibility Results of Different Grading Intervals r and the Highest Correlation Degree of Environmental Factors

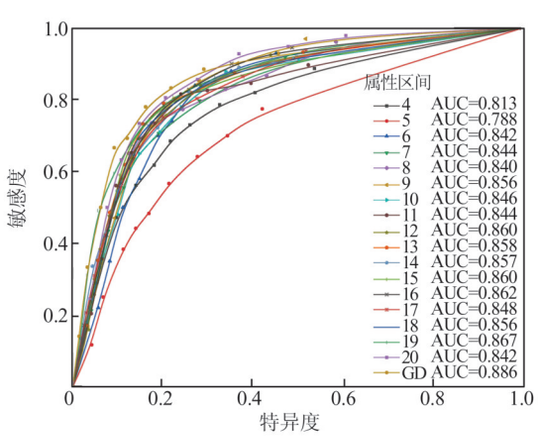


图 7　不同易发性结果对应的 ROC 曲线

Fig. 7　ROC Curves Corresponding to Different Susceptibility Results

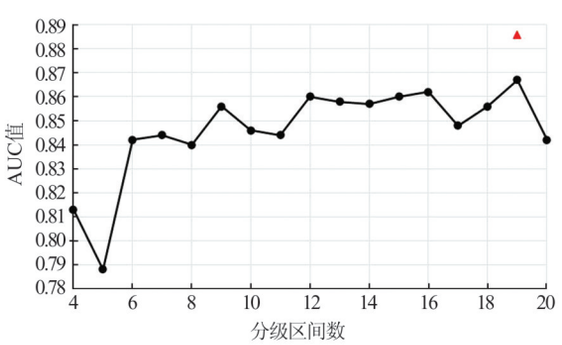


图 8　不同分级区间数与易发性结果的关系

Fig. 8　Relationship Between Numbers of Different Grading Interval and Susceptibility Results

分级区间数和 AUC 值是正相关的关系。在本研究中,并未出现明显的分界点,而且分级区间数和 AUC 值也不是正相关的关系。从图 8 可以看出,随着分级区间数的增加,AUC 值有增有减,单从 4、8、12、16、20 这 5 个点来看,AUC 值在分级区间数从 4 增大到 16 时是上升的,但是在分级区间数从 16 增大到 20 时 AUC 值却是下降的。出现这种情况的原因是地质灾害的发育规律具

有区域的独特性,不同区域的地质条件不同,形成灾害的条件各不相同,而本次讨论的通过计算分级区间与灾害点的关联度确定连续变量因子最优分级区间的方法可以推广到不同区域的地质灾害易发性研究。

与文献[14]的研究不同的是,本文将连续变量因子分级区间数划分为多种情况,分别计算其对易发性的影响,用GD计算得出最优分级区间。而文献[14]研究的是通过灾害点的数据提取灾害区域内的非连续变量因子,然后用自然间断点法对灾害区域内的非连续变量因子数据进行分类,这样虽然把灾害点和环境因子联系的较为紧密,但是由于一个区域内的灾害点数量有限,灾害区域内的环境因子数据较少,再加上灾害编录的人工误差的影响,这种分类方法只适用于小区域内灾害信息较为明确的易发性计算。此次研究了全区域的环境因子数据,利用全区域数据也会减弱环境因子与灾害的关联度,为了解决此问题,设置了多种环境因子区间分级情况,并采用GD计算不同环境因子的不同区间与灾害点的关联度,筛选不同环境因子与滑坡关联度最高的分级区间,通过易发性验算,证明这种环境因子的区间分级是最优的。

5 不同机器学习模型的结果对比研究

5.1 易发性评价结果

通过BPNN、XGBoost、RF 3种机器学习模型,采用半监督学习的方法[43],从研究区的非滑坡区域随机选取与滑坡点相等的非滑坡数据,将选取的非滑坡数据和滑坡点数据作为训练集训练模型,再将研究区的全部数据代入到训练好的模型计算易发性。将环境因子最优分级区间数据和滑坡灾害数据代入模型中计算研究区易发性结果如图9所示。

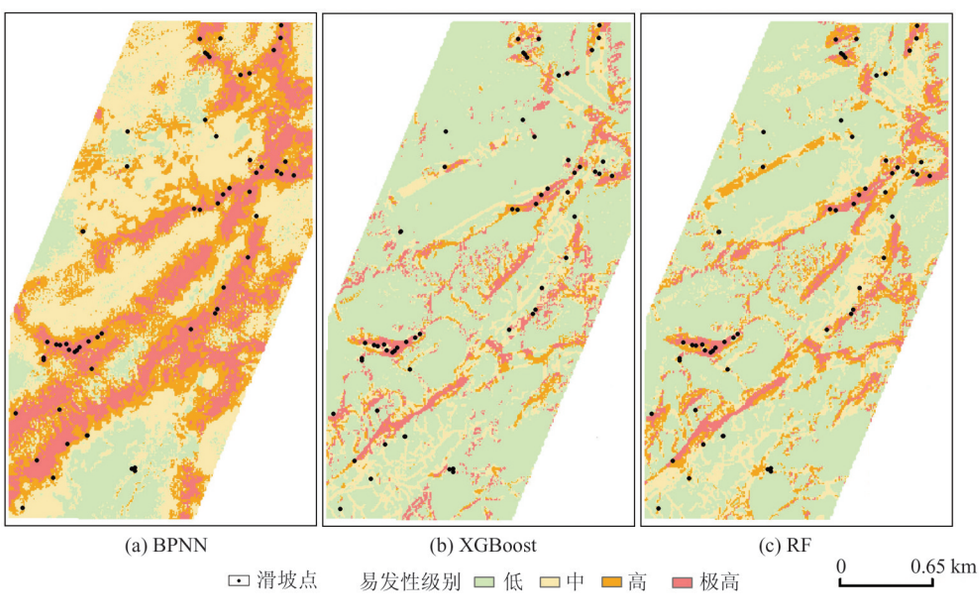


图9 3种模型计算易发性结果图

Fig. 9 Susceptibility Results of Three Models

5.2 精度统计分析

根据3种模型计算得到的4个等级分区滑坡的分布比例和研究区域分布比例如表4所示。由表4中数据分析可得:

1)BPNN模型中有53.33%的滑坡位于极高易发区,随着易发性等级的降低,易发区内的滑坡数量在逐渐减小,但是低易发区的面积较小,只占研究区总面积的12.88%,BPNN模型计算的结果把研究区更多地划分为中易发区,这与易发性分区原则是相悖的;

2)XGBoost模型中有63.33%的滑坡位于极高易发区,随着易发性等级的降低,易发区内的滑坡数量在逐渐减小,高易发区的面积低于极高易发区,较不符合易发性分区原则;

3)RF模型中有68.33%的滑坡位于极高易发区,随着易发性等级的降低,易发区内的滑坡数量在逐渐减小,分区面积逐渐增大,频率比逐渐减小,在理论上更符合易发性分区原则。

综上可知,BPNN模型的精度最差,不仅极高易发区包含的滑坡点较少,划分的研究区易发性等级也不符合易发性分区规律;XGBoost模型的精度次之,而RF模型的精度最高,极高易发区

表 4　3种模型易发性等级统计表

Table 4　Susceptibility Levels of Three Models

模型	易发性分区等级	发生滑坡数量	分区面积/km^2	占总滑坡比例/%	占总面积比例/%	频率比
BPNN	极高	32	24.93	53.33	19.16	2.78
	高	21	34.13	35.00	26.23	1.33
	中	7	54.29	11.67	41.73	0.28
	低	0	16.76	0	12.88	0
XGBoost	极高	38	11.14	63.33	8.56	7.39
	高	9	7.12	15.00	5.47	2.74
	中	13	18.48	21.67	14.21	1.53
	低	0	93.37	0	71.76	0
RF	极高	41	10.03	68.33	7.71	8.86
	高	18	13.80	30.00	10.61	2.83
	中	1	24.94	1.67	19.17	0.09
	低	0	81.34	0	62.51	0

包含的滑坡点最多,且划分的易发性等级符合易发性分区规律。

5.3　ROC 曲线分析

为了对比 BPNN 模型、XGBoost 模型和 RF 模型计算的易发性结果,画出 ROC 曲线如图 10 所示。经过计算可知,RF 模型 AUC 值最高为 0.886,相较于 BPNN 模型精度提高了 9.7%,相较于 XGBoost 模型精度提高了 9.6%。虽然 RF 模型和 XGBoost 模型都属于分类树模型,但是 RF 模型的表现更好,说明 RF 模型是这 3 种模型中最适合该研究区易发性计算的模型。

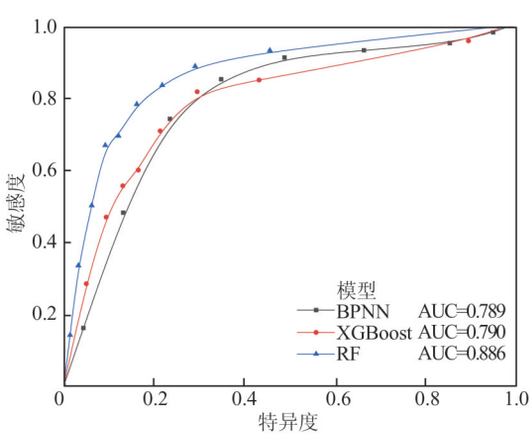


图 10　3种模型 ROC 对比图

Fig. 10　ROC Comparison of Three Models

6　结　语

本文通过以赣南地区宁都县赖村镇为例,基于主客观组合赋权法筛选环境因子,通过利用 GD 计算不同分级区间的连续变量因子与灾害点的关联度,从而选取最优分级区间,再分别采用 3 种不同模型开展滑坡易发性研究,得到结论如下:

1)通过 GD 得到的与灾害关联度最高的分区组合计算出的最大 AUC 值为 0.886,说明该方法可以得到最优的分级区间,且能有效提高易发性评价结果的精度。

2)通过对比 3 种模型的精度,RF 模型计算结果更加准确,对高易发性区域和低易发性区域的分类和预测更加科学;ROC 分析显示 RF 模型的 AUC 值为 0.886,BPNN 模型和 XGBoost 模型的 AUC 值分别为 0.789、0.790,相较于 BPNN 和 XGBoot 这两种模型,RF 预测精度分别提高了 9.7%、9.6%。

参　考　文　献

[1] PETLEY D. Global Patterns of Loss of Life from Landslides[J]. *Geology*, 2012, 40(10): 927-930.

[2] 徐胜华, 刘纪平, 王想红, 等. 熵指数融入支持向量机的滑坡灾害易发性评价方法: 以陕西省为例[J]. 武汉大学学报(信息科学版), 2020, 45(8): 1214-1222.
XU Shenghua, LIU Jiping, WANG Xianghong, et al. Landslide Susceptibility Assessment Method Incorporating Index of Entropy Based on Support Vector Machine: A Case Study of Shaanxi Province [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(8): 1214-1222.

[3] CHEN Y, DONG J L, GUO F, et al. Review of Landslide Susceptibility Assessment Based on Knowledge Mapping[J]. *Stochastic Environmental Research and Risk Assessment*, 2022, 36(9): 2399-2417.

[4] 闫举生, 谭建民. 基于不同因子分级法的滑坡易发性评价: 以湖北远安县为例[J]. 中国地质灾害与

防治学报, 2019, 30(1): 52-60.
YAN Jusheng, TAN Jianmin. Landslide Susceptibility Assessment Based on Different Factor Classification Methods—A Case Study in Yuan'an County of Hubei Province[J]. *The Chinese Journal of Geological Hazard and Control*, 2019, 30(1): 52-60.

[5] 李萍, 叶辉, 谈树成. 基于层次分析法的永德县地质灾害易发性评价[J]. 水土保持研究, 2021, 28(5): 394-399.
LI Ping, YE Hui, TAN Shucheng. Evaluation of Geological Hazards in Yongde County Based on Analytic Hierarchy Process[J]. *Research of Soil and Water Conservation*, 2021, 28(5): 394-399.

[6] 连志鹏, 徐勇, 付圣, 等. 采用多模型融合方法评价滑坡灾害易发性: 以湖北省五峰县为例[J]. 地质科技通报, 2020, 39(3): 178-186.
LIAN Zhipeng, XU Yong, FU Sheng, et al. Landslide Susceptibility Assessment Based on Multi-Model Fusion Method: A Case Study in Wufeng County, Hubei Province[J]. *Bulletin of Geological Science and Technology*, 2020, 39(3): 178-186.

[7] 吴辰文, 梁靖涵, 王伟, 等. 基于递归特征消除方法的随机森林算法[J]. 统计与决策, 2017, 33(21): 60-63.
WU Chenwen, LIANG Jinghan, WANG Wei, et al. Random Forest Algorithm Based on Recursive Feature Elimination[J]. *Statistics & Decision*, 2017, 33(21): 60-63.

[8] 游士兵, 严研. 逐步回归分析法及其应用[J]. 统计与决策, 2017, 33(14): 31-35.
YOU Shibing, YAN Yan. Stepwise Regression Analysis and Its Application[J]. *Statistics & Decision*, 2017, 33(14): 31-35.

[9] LUO W, LIU C C. Innovative Landslide Susceptibility Mapping Supported by Geomorphon and Geographical Detector Methods[J]. *Landslides*, 2018, 15(3): 465-474.

[10] 刘婷, 谭建民, 郭飞, 等. 人工切坡下滑坡易发性评价中权重修正方法研究: 以赣州市沙地镇为例[J]. 自然灾害学报, 2021, 30(5): 217-225.
LIU Ting, TAN Jianmin, GUO Fei, et al. Research on the Method of Weight Correction for Landslide Susceptibility with Artificial Cutting Slope: A Case Study of Shadi Town, Ganzhou City[J]. *Journal of Natural Disasters*, 2021, 30(5): 217-225.

[11] CHANG M, ZHOU Y, ZHOU C, et al. Coseismic Landslides Induced by the 2018 Mw 6.6 Iburi, Japan, Earthquake: Spatial Distribution, Key Factors Weight, and Susceptibility Regionalization[J]. *Landslides*, 2021, 18(2): 755-772.

[12] HUANG F M, YE Z, JIANG S H, et al. Uncertainty Study of Landslide Susceptibility Prediction Considering the Different Attribute Interval Numbers of Environmental Factors and Different Data-Based Models[J]. *CATENA*, 2021, 202: 105250.

[13] 凌晓, 刘甲美, 王涛, 等. 基于致灾因子对称法分级的信息量模型在地震滑坡危险性评价中的应用[J]. 国土资源遥感, 2021, 33(2): 172-181.
LING Xiao, LIU Jiamei, WANG Tao, et al. Application of Information Value Model Based on Symmetrical Factors Classification Method in Landslide Hazard Assessment[J]. *Remote Sensing for Land & Resources*, 2021, 33(2): 172-181.

[14] 孙德亮, 马祥龙, 唐小娅, 等. 基于不同因子分级的滑坡易发性区划对比: 以万州区为例[J]. 重庆师范大学学报(自然科学版), 2021, 38(5): 43-54.
SUN Deliang, MA Xianglong, TANG Xiaoya, et al. Comparison of Landslide Susceptibility Mapping Based on Different Factor Classifications: Taking Wanzhou District as an Example[J]. *Journal of Chongqing Normal University (Natural Science)*, 2021, 38(5): 43-54.

[15] 孙德亮. 基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D]. 上海: 华东师范大学, 2019.
SUN Deliang. Landslide Susceptibility Zoning and Rainfall-Induced Landslide Prediction and Early Warning Based on Machine Learning[D]. Shanghai: East China Normal University, 2019.

[16] YOUSSEF A M, PRADHAN B, DIKSHIT A, et al. Landslide Susceptibility Mapping Using CNN-1D and 2D Deep Learning Algorithms: Comparison of Their Performance at Asir Region, KSA[J]. *Bulletin of Engineering Geology and the Environment*, 2022, 81(4): 165-176.

[17] 赵占莺, 王继周, 毛曦, 等. 多维CNN耦合的滑坡易发性评价方法[J]. 武汉大学学报(信息科学版), 2024, 49(8): 1466-1481.
ZHAO Zhan'ao, WANG Jizhou, MAO Xi, et al. A Multi-dimensional CNN Coupled Landslide Susceptibility Assessment Method[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(8): 1466-1481.

[18] 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.

[19] 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.

[20] 黄发明, 李金凤, 王俊宇, 等. 考虑线状环境因子适宜性和不同机器学习模型的滑坡易发性预测建模规律[J]. 地质科技通报, 2022, 41(2): 44-59.
HUANG Faming, LI Jinfeng, WANG Junyu, et al. Modelling Rules of Landslide Susceptibility Prediction Considering the Suitability of Linear Environmental Factors and Different Machine Learning Models [J]. *Bulletin of Geological Science and Technology*, 2022, 41(2): 44-59.

[21] 赵久彬, 刘元雪, 刘娜, 等. 海量监测数据下分布式BP神经网络区域滑坡空间预测方法[J]. 岩土力学, 2019, 40(7): 2866-2872.
ZHAO Jiubin, LIU Yuanxue, LIU Na, et al. Spatial Prediction Method of Regional Landslide Based on Distributed BP Neural Network Algorithm Under Massive Monitoring Data[J]. *Rock and Soil Mechanics*, 2019, 40(7): 2866-2872.

[22] 田乃满, 兰恒星, 伍宇明, 等. 人工神经网络和决策树模型在滑坡易发性分析中的性能对比[J]. 地球信息科学学报, 2020, 22(12): 2304-2316.
TIAN Naiman, LAN Hengxing, WU Yuming, et al. Performance Comparison of BP Artificial Neural Network and CART Decision Tree Model in Landslide Susceptibility Prediction[J]. *Journal of Geo-Information Science*, 2020, 22(12): 2304-2316.

[23] 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.

[24] ZHOU X Z, WEN H J, LI Z W, et al. An Interpretable Model for the Susceptibility of Rainfall-Induced Shallow Landslides Based on SHAP and XGBoost[J]. *Geocarto International*, 2022, 37(26): 13419-13450.

[25] 刘艳辉, 黄俊宝, 肖锐铧, 等. 基于随机森林的福建省区域滑坡灾害预警模型研究[J]. 工程地质学报, 2022, 30(3): 944-955.
LIU Yanhui, HUANG Junbao, XIAO Ruihua, et al. Study on Early Warning Model for Regional Landslides Based on Random Forest in Fujian Province [J]. *Journal of Engineering Geology*, 2022, 30(3): 944-955.

[26] 仉文岗, 何昱苇, 王鲁琦, 等. 基于水系分区的滑坡易发性机器学习分析方法: 以重庆市奉节县为例[J]. 地球科学, 2023, 48(5): 2024-2038.
ZHANG Wengang, HE Yuwei, WANG Luqi, et al. Machine Learning Solution for Landslide Susceptibility Based on Hydrographic Division: Case Study of Fengjie County in Chongqing[J]. *Earth Science*, 2023, 48(5): 2024-2038.

[27] KONG C F, TIAN Y P, MA X G, et al. Landslide Susceptibility Assessment Based on Different Machine Learning Methods in Zhaoping County of Eastern Guangxi [J]. *Remote Sensing*, 2021, 13(18): 3573.

[28] JENNIFER J J. Feature Elimination and Comparison of Machine Learning Algorithms in Landslide Susceptibility Mapping [J]. *Environmental Earth Sciences*, 2022, 81(20): 489.

[29] 郭飞, 蒋广辉, 黄晓虎, 等. 环境因子组合和负样本选取策略对花岗岩区崩岗易发性评价的影响[J]. 农业工程学报, 2024, 40(1): 191-200.
GUO Fei, JIANG Guanghui, HUANG Xiaohu, et al. Impact of Environmental Factor Combinations and Negative Sample Selection on Benggang Susceptibility Assessment in Granite Areas [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(1): 191-200.

[30] 郭飞, 赖鹏, 黄发明, 等. 基于知识图谱的滑坡易发性评价文献综述及研究进展[J]. 地球科学, 2024, 49(5): 1584-1606.
GUO Fei, LAI Peng, HUANG Faming, et al. Literature Review and Research Progress of Landslide Susceptibility Mapping Based on Knowledge Graph [J]. *Earth Science*, 2024, 49(5): 1584-1606.

[31] 赵晓燕, 谈树成, 李永平. 基于斜坡单元与组合赋权法的东川区地质灾害危险性评价[J]. 云南大学学报(自然科学版), 2021, 43(2): 299-305.
ZHAO Xiaoyan, TAN Shucheng, LI Yongping. Risk Assessment of Geological Hazards in Dongchuan District Based on the Methods of Slope Unit and Combination Weighting[J]. *Journal of Yunnan University* (*Natural Sciences Edition*), 2021, 43(2): 299-305.

[32] 王劲峰, 徐成东. 地理探测器: 原理与展望[J]. 地理学报, 2017, 72(1): 116-134.
WANG Jinfeng, XU Chengdong. GeoDetector:

Principle and Prospective[J]. *Acta Geographica Sinica*, 2017, 72(1): 116-134.

[33] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Representations by Back-Propagating Errors[J]. *Nature*, 1986, 323: 533-536.

[34] CHEN T Q, GUESTRIN C. XGBoost: A Scalable Tree Boosting System[C]//The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016.

[35] BREIMAN L. Random Forests[J]. *Machine Learning*, 2001, 45: 5-32.

[36] 陈玉萍, 袁志强, 周博, 等. 遗传算法优化BP网络在滑坡灾害预测中的应用研究[J]. 水文地质工程地质, 2012, 39(1): 114-119.
CHEN Yuping, YUAN Zhiqiang, ZHOU Bo, et al. Application of back Propagation Neural Networks with Optimization of Genetic Algorithms to Landslide Hazard Prediction[J]. *Hydrogeology & Engineering Geology*, 2012, 39(1): 114-119.

[37] 刘坚, 李树林, 陈涛. 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), 2018, 43(7): 1085-1091.
LIU Jian, LI Shulin, CHEN Tao. Landslide Susceptibility Assesment Based on Optimized Random Forest Model[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(7): 1085-1091.

[38] 洪增林, 李永红, 张玲玉, 等. 一种基于主成分分析法的区域性地质灾害危险性评估方法[J]. 灾害学, 2020, 35(1): 118-124.
HONG Zenglin, LI Yonghong, ZHANG Lingyu, et al. A Regional Geological Hazard Risk Assessment Method Based on Principal Component Analysis[J]. *Disaster Science*, 2020,35(1): 118-124.

[39] 孟凡奇, 高峰, 林波, 等. 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, 2023, 38(3): 111-117.
MENG Fanqi, GAO Feng, LIN Bo, et al. Evaluation of Geological Disaster Susceptibility in Eastern Shandong Based on AHP and Information Model[J]. *Journal of Catastrophology*, 2023, 38(3): 111-117.

[40] 易思材, 张明文. 绿春县半坡乡某村泥石流发育特征及危险性评价[J]. 地质灾害与环境保护, 2022, 33(1): 15-23.
YI Sicai, ZHANG Mingwen. Development Characteristics and Risk Assessment of Baliu Debris Flow in Banpo Township of Luchun County[J]. *Journal of Geological Hazards and Environment Preservation*, 2022, 33(1): 15-23.

[41] 黄智杰, 简文彬, 夏昌, 等. 基于LSO-RF模型的阶跃型滑坡位移速率预测方法[J]. 福州大学学报(自然科学版), 2023, 51(6): 872-878.
HUANG Zhijie, JIAN Wenbin, XIA Chang, et al. Prediction of Displacement Rate of Step-Like Landslide Based on LSO-RF Model[J]. *Journal of Fuzhou University* (*Natural Science Edition*), 2023, 51(6): 872-878.

[42] 中华人民共和国自然资源部. DZ/T 0438—2023, 地质灾害风险调查评价规范(1:50 000)[S]. 北京: 自然资源部, 2023.
Ministry of Natural Resources of the People's Republic of China. DZ/T 0438—2023, Standard for Geohazard Risk Survey and Evaluation (1:50 000)[S]. Beijing: Ministry of Natural Resources, 2023.

[43] 黄发明, 潘李含, 姚池, 等. 基于半监督机器学习的滑坡易发性预测建模[J]. 浙江大学学报(工学版), 2021, 55(9): 1705-1713.
HUANG Faming, PAN Lihan, YAO Chi, et al. Landslide Susceptibility Prediction Modelling Based on Semi-Supervised Machine Learning[J]. *Journal of Zhejiang University* (*Engineering Science*), 2021, 55(9): 1705-1713.