

# 基于 GBDT-LR 和信息量模型耦合的滑坡易发性评价

董张玉<sup>1,2,3</sup>, 张晋<sup>1</sup>, 彭鹏<sup>4</sup>, 汪燕<sup>4</sup>, 杨智<sup>4</sup>, 安森<sup>1</sup>

[1.合肥工业大学 计算机与信息学院, 安徽 合肥 230601; 2.工业安全与应急技术安徽省重点实验室, 安徽 合肥 230601;  
3.智能互联系统安徽省实验室, 安徽 合肥 230601; 4.安徽省地质调查院(安徽省地质科学研究所), 安徽 合肥 230001]

**摘要:** [目的] 探索准确、快速的滑坡易发性区划方法, 为区域安全监测提供参考, 为政府治理滑坡灾害提供科学依据。[方法] 以安徽省池州市贵池区为研究区域, 采用梯度提升决策树一逻辑回归(GBDT-LR)和信息量(I)模型耦合的方法, 实现区域滑坡易发性评价。该方法通过对原样本地学习, 组合产生新的模拟样本, 从而增强易发性评价模型对滑坡的拟合能力; 采用 Borderline-Smote 算法解决样本数据不对称的问题。选用 r.slopeunits 软件划分的斜坡单元作为最小评价单元, 选取坡度、坡向、地形曲率、剖面曲率、平面曲率、地形湿度指数(TWD)、地形起伏度、归一化植被指数(NDVI)、距断裂距离和距水系距离总计 10 个评价因子。分别从频率比、滑坡灾害点及隐患点密度、ROC 曲线 3 个方面对构建的滑坡易发性模型进行评价。[结果] 试验结果表明, 耦合模型 I-GBDT-LR 分别比 I, LR, I-LR 模型的高易发区频率比所占比例提升约 10%, 13%, 7%, 高易发区滑坡灾害点及隐患点密度分别提升约 9, 11, 7, ROC 精度提升约 10%, 9%, 5%。[结论] 从检验指标综合来看, 耦合模型的精度均高于单一模型, 所提出耦合模型精度又高于 I-LR 耦合模型, 为滑坡易发性评价提供了一种有效的、新型的评价方法。

**关键词:** 滑坡易发性; 信息量; 逻辑回归; GBDT-LR; 安徽省池州市

**文献标识码:** B

**文章编号:** 1000-288X(2023)01-0149-09

**中图分类号:** P642.22

**文献参数:** 董张玉, 张晋, 彭鹏, 等. 基于 GBDT-LR 和信息量模型耦合的滑坡易发性评价[J]. 水土保持通报, 2023, 43(1): 149-157. DOI: 10.13961/j.cnki.stbctb.2023.01.018; Dong Zhangyu, Zhang Jin, Peng Peng, et al. Landslide susceptibility evaluation based on coupling of GBDT-LR model and information model [J]. Bulletin of Soil and Water Conservation, 2023, 43(1): 149-157.

## Landslide Susceptibility Evaluation Based on Coupling of GBDT-LR Model and Information Model

Dong Zhangyu<sup>1,2,3</sup>, Zhang Jin<sup>1</sup>, Peng Peng<sup>4</sup>, Wang Yan<sup>4</sup>, Yang Zhi<sup>4</sup>, An Sen<sup>1</sup>

[1.School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China; 2.Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230601, China; 3.Intelligent Interconnected Systems Laboratory of Anhui Provincial, Hefei, Anhui 230601, China; 4.Geological Survey Anhui Province (Anhui Institute of Geological Sciences), Hefei, Anhui 230001, China]

**Abstract:** [Objective] The accurate and rapid landslide susceptibility zoning method were studied in order to provide a reference for regional safety monitoring, and provide a scientific basis for the government to control landslide disasters. [Methods] The study was conducted in the Guichi District of Chizhou City, Anhui Province. The coupled model of gradient boosting decision tree-logistic regression (GBDT-LR) and an information value (I) model was used to determine the evaluation of regional landslide susceptibility. The model learns from the original samples and combines them to generate new simulation samples in order to enhance the fitting ability of the model to evaluate landslide susceptibility. The Borderline-Smote algorithm was used to solve the problem of sample data asymmetry. The slope unit divided by r.slopeunits software was selected as the minimum evaluation unit, and a total of 10 evaluation factors were selected: slope gradient,

收稿日期: 2022-04-09

修回日期: 2022-06-27

资助项目: 安徽省重点研究与开发计划项目“典型地质灾害天地一体化协同监测关键技术研究”(20200407020030); 中央高校基本科研业务费专(JZ2021HG TB0111); 安徽省自然科学基金资助项目(2108085MF233)

第一作者: 董张玉(1986—), 男(汉族), 安徽省安庆市人, 博士, 教授, 硕士生导师, 主要从事高分遥感空间信息处理及其应用。Email: dzyhfut@hfut.edu.cn.

slope aspect, terrain curvature, profile curvature, plane curvature, topographic wetness index (TWI), topographic relief, normalized difference vegetation index (NDVI), distance from fault, and distance from river. The landslide susceptibility model was evaluated from three aspects: frequency ratio, density of landslide disaster points and hidden danger points, and the receiver operating characteristic (ROC) curve. [Results] The experimental results showed that the frequency ratio of the coupled model I-GBDT-LR was 10%, 13%, and 7% greater than that of the I, LR, and I-LR models, respectively. The density of landslide disaster points and hidden danger points in the high risk area increased by about 9, 11, and 7, respectively, and the ROC accuracy increased by about 10%, 9%, and 5%, respectively. [Conclusion] The accuracy of the coupled model was higher than that of the single model, and the accuracy of the coupled model proposed was higher than that of the I-LR coupled model, which provides an effective and new evaluation method for landslide susceptibility evaluation.

**Keywords: landslide susceptibility; information value; logistic regression; gradient boosting decision tree-logistic regression (GBDT-LR); Chizhou City of Anhui Province**

滑坡、崩塌和泥石流是中国典型的 3 种地质灾害,并且中国是世界上受滑坡影响最严重的国家之一,滑坡作为最常见的地质灾害类型,其特点是分布范围广,发生频率高,移动速度快,破坏性大等<sup>[1]</sup>。滑坡易发性评价是滑坡预测研究方面的重点,通过研究区域的遥感数据,提取滑坡各个方面(地形地貌、地层岩性、水文等)的影响因子,定量分析影响因子权重,进而预测滑坡在研究区域的空间分布和发生概率,从而为管理部门开展滑坡防治工作提供一定的数据支撑,提前对滑坡发生可能性较大的县镇村庄做好预防措施,减少人员和财产损失,提高滑坡地质灾害的治理效率<sup>[2-4]</sup>。

滑坡易发性评价的关键问题在于评价模型的选取,评价模型选取的合理与否,直接影响最终滑坡易发性评价的准确性。针对此问题,国内外学者在进行滑坡易发性评价时选取了各种各样的模型。目前,国内外学者通常采用逻辑回归<sup>[5-6]</sup>、信息量<sup>[7-8]</sup>、层次分析<sup>[9-10]</sup>、支持向量机<sup>[11]</sup>、机器学习<sup>[12-13]</sup>、深度学习等<sup>[14]</sup>模型的研究方法。例如,国内学者,罗路广等<sup>[5]</sup>采用逻辑回归和 K 折交叉验证法,对九寨沟地区进行滑坡易发性评价和因子重要性进行排序;许英姿等<sup>[7]</sup>采用 GIS 技术和信息量模型,对广西花岗岩分布区进行滑坡易发性评价;申怀飞等<sup>[9]</sup>采用层次分析法与信息量相结合的方法,对甘肃省进行滑坡易发性评价;徐胜华等<sup>[11]</sup>采用熵指数、支持向量机再结合 PSO 算法的方法,对陕西省进行滑坡易发生评价;黄发明等<sup>[13]</sup>采用半监督卡方自交互侦测决策树和半监督反向传播神经网络,对江西省南康区进行滑坡易发性评价;王世宝等<sup>[14]</sup>采用深度学习中的卷积神经网络,对川藏铁路康定—理塘段进行滑坡易发性评价,并与人工神经网络进行对比分析;国外学者, Ali Polat

等<sup>[15]</sup>采用频率比、信息量、逻辑回归、随机森林和多层感知机模型,对土耳其的西瓦斯市进行滑坡敏感性预测;Hung<sup>[16]</sup>采用层次分析法和加权线性组合法相结合的方法,对越南北部罗江上游流域进行滑坡易发性评价;Zhice Fang 等<sup>[17]</sup>以支持向量机为基础,对三峡大坝秭归—巴东段进行滑坡易发性评价,并对对比分析了深度学习及支持向量机在不同核函数情况下的评价结果;Somnath Bera 等<sup>[18]</sup>采用深度学习模型,对印度喜马拉雅山进行滑坡敏感性建模。总结得出,基于斜坡单元的滑坡易发性评价一般步骤是先提取研究区的评价因子,再对评价因子进行分级,最后确定合适的模型。

在滑坡易发性评价中,由于致灾原因的复杂性,单独使用信息量模型相当于默认各影响因子对滑坡影响的权重相等,影响因子之间的差异性被屏蔽<sup>[19]</sup>。而逻辑回归模型容易理解、训练速度快、不需要对评价因子进行缩放、可以得到各个影响因子权重,可以弥补信息量模型的不足。因此,当前滑坡易发性评价中,将信息量模型和逻辑回归模型相结合的方式得到了广泛应用,但是又引入了新的问题:由于逻辑回归的局限性,通过逻辑回归对影响因子赋予权重的方法,只对线性模型具有较好的拟合性,对非线性模型拟合能力较差<sup>[16]</sup>,即逻辑回归只是对样本数据进行表层分析,无法深入挖掘出影响因子的非线性特征。然而,在滑坡发育的实际过程中,受多种地形地貌、地质等因子影响,因子之间的逻辑关系十分复杂,仅靠信息量和逻辑回归模型简单相结合的方式,很难揭示出其内在的关系<sup>[20]</sup>。为了解决这个问题,可以在信息量(I)和逻辑回归(LR)结合的过程中加入梯度提升决策树(GBDT),因为 GBDT 可以完成特征从低维度到高维度的转换成,实则就是将滑坡样本从低维特征转换成高维特征,这样就可以将数据中的非线性特

征转换成线性特征,再将转换后的结果输入到 LR 中进行拟合,理论上将会得到更好滑坡预测效果。

基于上述想法,论文以池州市贵池区为研究区域,结合研究区域滑坡发育的特点,综合考虑前人研究成果<sup>[5,9,21-22]</sup>,最终确定 10 个评价因子,基于信息量模型(I)、梯度提升决策树—逻辑回归(GBDT-LR)算法构建滑坡易发性评价模型,采用 Borderline-Smote 算法解决滑坡样本的不对称问题,将信息量结果从低维度转换成高维度,再使用 LR 进行拟合。通过频率比、滑坡灾害点及隐患点密度和 ROC 曲线对评价模型进行评估,最终得到贵池区易发性区划,并结合实际情况分析所制成的滑坡易发性评价结果的合理性。以期为贵池区滑坡发生的可能性提供参考,为政府治理滑坡灾害提供科学的参考依据。

## 1 研究区概况与数据来源

### 1.1 研究区概况

研究区域为位于安徽省南部的贵池区,介于东经 117°06′—117°50′,北纬 30°15′—30°48′之间,全区面积约 2 516 km<sup>2</sup>。北接长江,南至黄山、九华山,东北、西北分别与铜陵、安庆毗邻,西南、东南分别与本市东至县、石台县和青阳县相接。贵池区属北亚热带湿润性季风气候区,气候温顺,降雨量适宜,光照充裕,四季分明,地表径流顺畅,水源发达;贵池区地处皖南山地与沿江丘陵平原过渡地带,其东南部多为山地,中部以丘陵为主,西北部为平原,整个地势自东南向西北倾斜。九华山西延余脉蜿蜒至贵池区东南,山岭纵横,峰峦绵延,主要山峰海拔高程多在 300~1 000 m,最高峰金家山海拔 1 025 m。

### 1.2 数据来源

本文研究贵池区滑坡易发性评价的主要数据为 DEM、断裂、归一化植被指数(NDVI)、水系以及滑坡点矢量数据。DEM、断裂、滑坡点数据均来源于当地有关部门,其中 DEM 分辨率为 15 m,滑坡点数据包含 86 个已经发生的历史滑坡点及 75 个滑坡灾害隐患点;NDVI 是使用 Landsat 8 多光谱卫星影像提取得到;水系数据来源于 OpenStreetMap 网站的开源矢量数据;试验中的数据采集时间均在 2021 年 5—8 月。由于数据来源不同,为了尽可能减少试验误差,在试验之前将所有数据进行预处理,预处理包括:地理坐标系变换、投影变换、图像配准、重采样、栅格矢量化、矢量栅格化等、几何校正等。

### 1.3 评价单元的划分

在 GRASS GIS 中,采用 r.slopeunits<sup>[23]</sup>斜坡单

元自动划分方法,通过不断调试,最终将研究区域划分为  $1.795 \times 10^4$  个斜坡单元。

## 2 研究方法

### 2.1 滑坡易发性评价模型的构建

2.1.1 信息量(I)模型 滑坡的发生受不同大小、不同性质的评价因子的影响,因此,利用信息量模型评价滑坡易发性的思路是:根据已有评价因子的实际测量值,与滑坡数据结合,将评价因子转化成反映实际滑坡的信息量值<sup>[19]</sup>。滑坡易发性评价是否准确与评价过程中获取的信息数量和质量密切相关,在实际试验过程中,常用滑坡样本频率计算信息量,如式(1)所示:

$$I = \sum_{i=1}^n I(x_i, Y) = \sum_{i=1}^n \ln \frac{N_i/N}{S_i/S} \quad (1)$$

式中: $I$  表示评价单元中总信息量; $n$  表示评级因子数量; $x_i$  表示当前评价单元内所取的评价因子等级; $Y$  表示滑坡事件; $S$  表示研究区域总面积; $S_i$  表示研究区域内含有  $x_i$  评价因子的总面积; $N$  表示研究区域内滑坡总面积; $N_i$  表示研究区域内含有  $x_i$  评价因子的滑坡总面积。信息量  $I$  作为滑坡评价的综合性指标,其值越大,滑坡发生的可能性越大,反之,发生滑坡的可能性越小。

2.1.2 逻辑回归(LR)模型 逻辑回归(Logistic Regression)模型是线性回归模型的扩展,也是经典的多元统计方法。逻辑回归模型被广泛应用于滑坡易发性评价建模中,可以较准确地揭示二元因变量(1 表示发生滑坡,0 表示未发生滑坡)和自变量(评价因子)之间的逻辑关系<sup>[24]</sup>。其函数表达式如下:

$$\begin{cases} \ln \frac{p}{1-p} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \\ P = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} = \frac{1}{1 + e^{-(\alpha + \beta x)}} \end{cases} \quad (2)$$

式中: $\alpha$  代表逻辑回归的常数项; $\beta$  代表逻辑回归的回归系数; $P$  代表滑坡发生概率, $P$  的值越接近 1,表明该斜坡单元发生滑坡的可能性越大,否则,表明发生滑坡的可能性越小。

2.1.3 梯度提升决策树(GBDT)模型 GBDT(gradient boosting decision tree)是提升算法 Boosting 的一种,其原理是将 gradient boosting 应用到决策树回归算法迭代的过程中,它主要包括三部分:回归树、梯度迭代、缩减。利用原始数据生成第一棵树后,以选取当前最小的损函数为目标生成第二棵树,依次迭代,直至损失函数残差达到预定值或达到最大迭代次数<sup>[25-26]</sup>。GBDT 算法过程如下:



假设输入训练数据集  $T = \{(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)\}$ ,  $x_i \in X \subseteq \mathbb{R}^n$ ,  $y_i \in y \subseteq \mathbb{R}$ ; 损失函数  $L[y, f(x)]$ ; 输出回归树  $\hat{f}(x)$ 。

(1) 初始化。

$$f_0(x) = \arg \min_c \sum_{i=1}^N L(y_i, c) \quad (3)$$

(2) 对  $m=1, 2 \dots M$ , 计算。

(a) 对  $i=1, 2 \dots N$ , 计算

$$r_{mi} = - \left\{ \frac{\partial L[y_i, f(x_i)]}{\partial f(x_i)} \right\} \quad f(x) = f_{m-1}(x) \quad (4)$$

(b) 对  $r_{mi}$  拟合一个回归树, 得到第  $m$  棵树的叶结点区域  $R_{mj}$ ,  $j=1, 2 \dots J$ 。

(c) 对  $j=1, 2 \dots J$ , 计算

$$c_{mj} = \arg \min_c \sum_{x_i \in R_{mj}} L[y_i, f_{m-1}(x_i) + c] \quad (5)$$

(d)  $\theta$  表示落入当前节点的样本集合, 更新

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J c_{mj} \theta(x \in R_{mj}) \quad (6)$$

(3) 得到回归树。

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} \theta(x \in R_{mj}) \quad (7)$$

2.1.4 I 与 GBDT-LR 模型耦合 鉴于 LR 模型的低复杂度和优秀的并行化处理能力的特点, 被广泛应用于滑坡易发性评价中, 并且取得了良好的效果。但是, LR 模型实质只是线性模型的泛化, 所以在拟合非线性模型时性能较差, 表达能力有限, 无法深入挖掘滑坡样本的非线性特征。因此, 需要对样本进行人工特征组合, 将非线性特征转换成线性特征, 再提供给 LR 学习。其存在的问题是, 在进行特征组合时, 需要专家的指导, 这极大地增加了数据的处理成本和难度。

GBDT 生成的每一棵树自根节点到叶节点都是由样本数据中的某些特征组合得到, 即一棵树代表一种特征组合。从公式(7)可知, 每棵树都有权重, 因此每种特征组合都有权重, 权重越大, 表明当前特征组合对结果影响越大。为了从 GBDT 中获得有效的特征组合, 可以将每棵树的叶节点看成一个维度, 并记录下样本数据经过每棵树后的叶节点位置, 并进行 one-hot 编码, 所得新的样本数据即为低维稠密特征到高维稀疏特征转换后的数据。为了避免 GBDT 过拟合, 只将原数据的部分输入 GBDT 中生成新样本, 将剩下的部分与新样本混合, 输入到 LR 中。

综上所述, GBDT-LR 与 I 模型耦合的过程及原理如下: 将滑坡数据和影响因子经过 I 模型后, 变成低维稠密特征样本数据, 然后输入到 GBDT 模型, 经过上述过程后, 滑坡样本数据从低维度映射到高维

度, 并产生新的有效特征组合, 这时滑坡样本中低维度的非线性特征将会转换成高维度的线性特征, 再利用 LR 模型对线性模型强大的拟合能, 从而提高滑坡的预测精度(图 1)。GBDT 和 LR 结合的算法, 是由 Xinran He 等<sup>[27]</sup>在 2014 年提出来的, 当时在广告点击率预测方面取得了非常优秀的效果。

2.1.5 滑坡易发性评价流程 本文从非滑坡单元中随机抽取 1 000 个非滑坡单元作为样本, 为了保证抽取的非滑坡样本的可靠性, 只从距水系距离、距断裂距离大于 500 m 的非滑坡单元中抽取数据, 作为非滑坡样本。由于本文滑坡点数据较少, 为了解决试验中样本不对称的问题, 采用 Borderline-Smote 算法进行数据采样, 生成和非滑坡样本相同数量的滑坡样本。具体的滑坡易发性评价流程图见图 2, 试验结果采用频率比、滑坡灾害点及隐患点密度和 ROC 精度曲线进行检验。

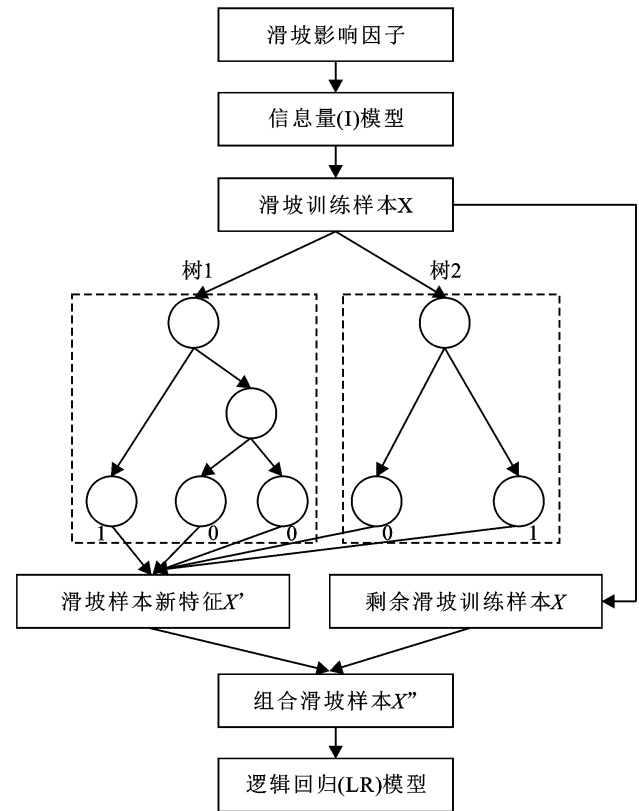


图 1 GBDT-LR 与 I 耦合过程

Fig.1 Coupling process of GBDT-LR and I

## 2.2 贵池区滑坡评价因子选取及信息量

2.2.1 评价因子的选取 滑坡的发生受多种环境因素的影响, 主要有地质因子、地质构造因子、地表覆被因子、地形地貌因子。合理地选取评价因子, 会大幅度提升滑坡灾害评价的准确度。本文通过对贵池区滑坡发生规律本质的总结, 再结合前人对滑坡易发性

评价经验<sup>[5,9,21-22]</sup>,最终选取 10 个评价因子,分别为:坡度、坡向、地形曲率、剖面曲率、平面曲率、地形湿度指数(TWI)、地形起伏度、归一化植被指数(NDVI)、

距断裂距离、距水系距离。其中地形地貌因子从 DEM 中提取,距断裂、水系距离是通过 ArcGIS 中的欧式距离分析得到。

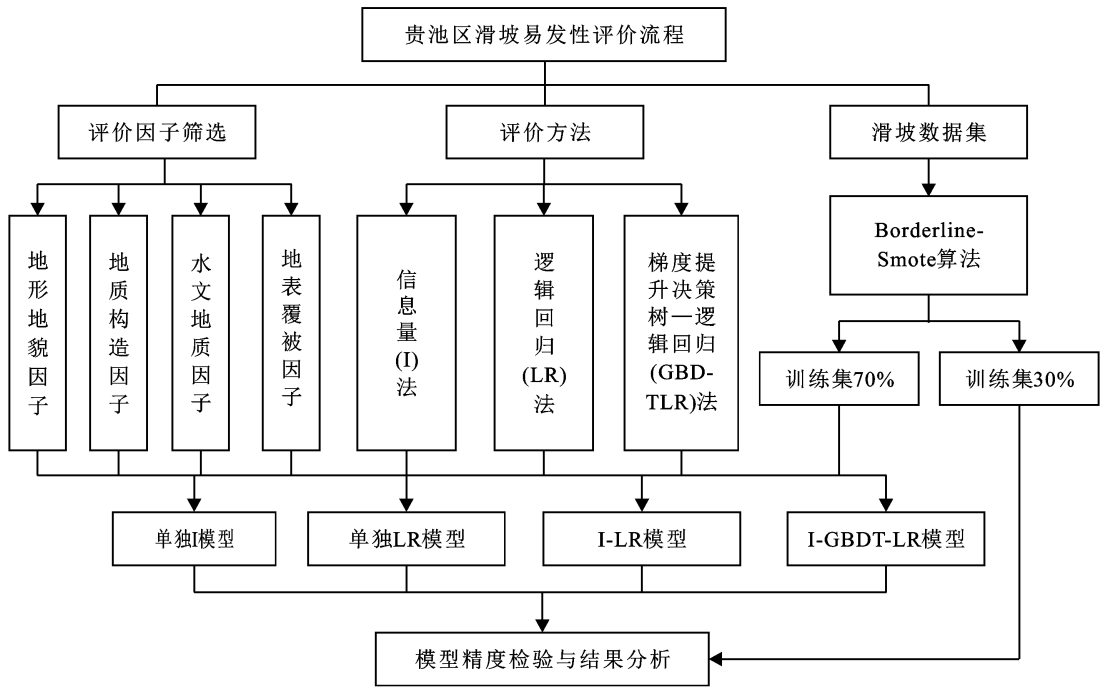


图 2 贵池区滑坡易发性评价流程

Fig.2 Evaluation process of landslide susceptibility in Guichi District

2.2.2 信息量值 在 ArcGIS 中,利用分区统计功能,对评价因子中连续类型的评价因子求平均值,例如某个斜坡单元的地形曲率值,是地形曲率因子与该斜坡单元对应位置内所有栅格地形曲率值的平均值;离散类型的评价因子求众数,例如某个斜坡单元的坡向,是坡向因子与该斜坡单元对应位置内所有栅格坡向值的众数;这样即可得到每个斜坡单元对应的评价因子值。然后,对评价因子进行分级,为了避免手动分级所造成的主观因素影响,论文参考 ArcGIS 中的自然断点法(Jenks' natural breaks)再结合前人经验,对评价因子进行分级。自然断点法是通过统计类别的方差,对相似值进行最恰当地分组,并可使各类之间的差异最大化。最后,利用频率比和信息量计算每个分级的频率比值和信息量值,结果见表 1。

### 3 结果与分析

#### 3.1 评价指标因子的共线性判断

在进行逻辑回归之前,需要对评价因子进行多重共线性检验,保证评价因子之间的独立性。如果评价因子之间存在多重共线性,不仅会增加模型复杂度,而且模型结果也不一定可靠<sup>[21]</sup>。因此,在逻辑回归模型中,多重共线性检验是必不可少的。本文采用容

忍度(tolerance, TOL)和方差膨胀因子(variance inflation factor, VIF)对 10 个评价因子进行多重共线性检验:

$$VIF = \frac{1}{TOL} = \frac{1}{1 - R_i^2} \quad (8)$$

式中: $R_i^2$ 表示第*i*个评价因子 $x_i$ 与其余评价因子之间的判定系数;当 $VIF > 10$ 或 $TOL < 0.1$ 表明该评价因子存在多重共线性,否则,不存在多重共线性。

本文采用 SPSS 软件进行多重共线性检验,检验结果见表 2。从表 2 中可以看出,所选所有评价因子 VIF 均小于 10,说明评价因子之间不存在多重共线性,因此,所有评价因子均可参与后续试验。

#### 3.2 评价模型的精度

ROC(receiver operation characteristic curves)曲线又称受试者工作特征曲线,在滑坡易发性模型精度评价中被广泛使用,它是通过 ROC 曲线下方的面积 AUC(Area Under ROC)进行评价,AUC 的值介于 0~1 之间,值越大说明模型分类效果越好,精度越高,否则越低<sup>[5,9,14,21]</sup>。在 ROC 曲线中横坐标代表假阳性率(FPR),其含义是非滑坡测试样本被预测为滑坡的个数;纵坐标代表真阳性率(TPR),其含义是滑坡测试样本中被正确预测的滑坡数。

表 1 研究区域滑坡灾害各评价因子指标分级以及信息量

Table 1 Studies the classification and information amount of each evaluation factor index of regional landslide disaster

评价因子	数据类型	因子分级	滑坡分级 面积/km <sup>2</sup>	滑坡面积 比例 A/%	区域分级 面积/km <sup>2</sup>	分级面积 比例 B/%	频率比 A/B	信息量 (I)
坡向	离散	平面	0	0	0.086	0.003	0.000	—
		北	3.330	16.129	317.324	12.667	1.273	0.242
		东北	1.409	6.827	243.805	9.732	0.701	-0.355
		东	1.261	6.108	322.995	12.893	0.474	-0.747
		东南	3.876	18.775	344.812	13.764	1.364	0.310
		南	3.203	15.514	287.835	11.490	1.350	0.300
		西南	1.940	9.398	249.881	9.975	0.942	-0.060
		西	2.398	11.613	320.491	12.793	0.908	-0.097
地形曲率	连续	西北	3.228	15.636	417.898	16.682	0.937	-0.065
		<-0.191~795	0.349	1.688	103.415	4.128	0.075	-0.894
		-0.192~-0.031	3.624	17.552	566.401	22.610	0.776	-0.253
		-0.031~0.131	14.905	72.197	1 648.177	65.792	1.097	0.093
剖面曲率	连续	>0.131	1.768	8.562	187.133	7.470	1.146	0.136
		<-0.050	0.499	2.418	112.384	4.486	0.539	-0.618
		-0.050~0.059	11.250	54.491	1 319.514	52.673	1.035	0.034
		0.059~0.172	7.897	38.250	895.488	35.746	1.070	0.068
平面曲率	连续	>0.172	1.000	4.841	177.740	7.095	0.682	-0.382
		<-0.070	0.445	2.157	121.415	4.847	0.445	-0.809
		-0.070~0.067	8.684	42.063	1 411.121	56.329	0.747	-0.292
		0.067~0.198	10.508	50.898	828.483	33.071	1.539	0.431
NDVI	连续	>0.198	1.008	4.881	144.107	5.752	0.848	-0.164
		<0.100	0.122	0.593	132.544	5.291	0.112	-2.189
		0.100~0.288	4.941	23.933	767.678	30.644	0.781	-0.247
距断裂距离	连续	>0.288	15.582	75.474	1 604.905	64.065	1.178	0.164
		<500 m	10.344	50.102	683.641	27.290	1.836	0.608
		500~1 500 m	5.845	28.310	830.512	33.152	0.854	-0.158
		1 500~2 500 m	2.326	11.264	340.726	13.601	0.828	-0.189
		2 500~3 500 m	0.678	3.282	146.241	5.838	0.562	-0.576
坡度	连续	>3500 m	1.454	7.041	504.006	20.119	0.350	-1.050
		<5°	1.510	7.313	763.106	30.462	0.240	-1.427
		5°~15°	2.733	13.237	505.414	20.175	0.656	-0.421
		15°~25°	8.236	39.893	580.329	23.166	1.722	0.544
		25°~35°	7.653	37.070	522.310	20.850	1.778	0.575
		35°~45°	0.513	2.486	132.866	5.304	0.469	-0.758
TWI	连续	>45°	0	0	1.101	0.043	0	—
		<5.5	16.294	78.923	1 264.589	50.480	1.563	0.447
		5.5~6.5	0.644	3.117	358.409	14.307	0.218	-1.524
		6.5~7.5	2.084	10.093	230.029	9.182	1.099	0.095
		7.5~8.5	1.397	6.764	174.827	6.979	0.969	-0.031
地形起伏度	连续	>8.5	0.228	1.102	477.272	19.052	0.058	-2.850
		<5 m	2.885	13.976	878.198	35.056	0.399	-0.920
		5~12 m	1.468	7.112	417.207	16.654	0.427	-0.851
		12~19 m	6.516	31.562	486.058	19.403	1.627	0.487
		19~26 m	7.004	33.924	445.733	17.793	1.907	0.645
距水系距离	连续	>26 m	2.772	13.425	277.931	11.094	1.210	0.191
		<1 068 m	5.903	28.592	652.239	26.036	1.098	0.094
		1 068~2 199 m	6.349	30.753	538.139	21.482	1.432	0.359
		2 199~3 424 m	2.160	10.462	488.054	19.482	0.537	-0.622
		3 424~4 860 m	2.673	12.946	348.467	13.910	0.931	-0.072
		4 860~6 656 m	0.890	4.313	247.928	9.897	0.436	-0.831
		6 656~8 944 m	2.524	12.227	147.044	5.870	2.083	0.734
>8 944 m	0.146	0.707	83.256	3.323	0.213	-1.547		

表 2 评价因子多重共线性检验结果

Table 2 Multicollinearity test results of evaluation factors

评价因子	TOL	VIF
坡向	0.923	1.084
地形曲率	0.798	1.254
剖面曲率	0.918	1.089
平面曲率	0.765	1.307
归一化植被指数(NDVI)	0.819	1.221
距断裂距离	0.699	1.431
坡度	0.197	5.073
地形湿度指数(TWI)	0.422	2.370
地形起伏度	0.234	4.274
距水系距离	0.895	1.117

根据图 3 可以看出,4 种评价模型的 ROC 曲线下的面积,即 AUC 值均大于 0.75,表明 4 种模型均能精确的对贵池区进行滑坡易发性评价。4 种评价模型的精度从小到大依次为 I, LR, I-LR, I-GBDT-LR, 其值分别为:84.58%,85.63%,90.13%,94.77%。且 I-LR 模型的精度与单独模型 I 和 LR 相比,精度提升约 5%,I-GBDT-LR 耦合模型比两个单一模型精度提升约 10%。

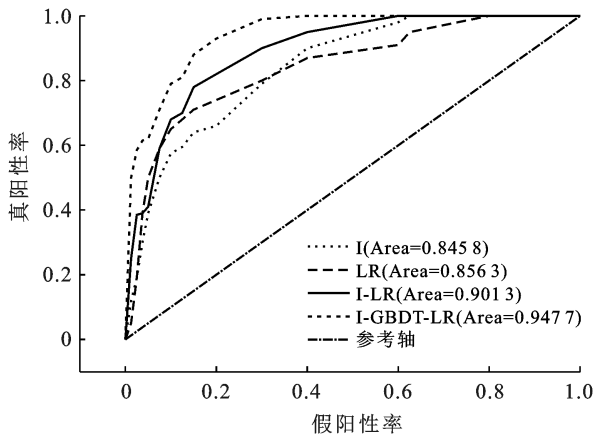


图 3 滑坡易发性评价模型的工作特征曲线(ROC)

Fig.3 Receive operating characteristic curve of landslide susceptibility evaluation model

### 3.3 贵池区易发性区划结果

利用 Python,将训练样本各斜坡单元评价因子的频率比或信息量进行归一化,输入到相应的模型中进行训练,得到贵池区滑坡易发性的概率分布。通过 GIS 中的自然断点法,对贵池区进行滑坡易发性等级划分,将其划分为 4 个等级,分别为:极低易发区、低易发区、中易发区和高易发区,再将贵池区滑坡易发性评价结果与已知的贵池区滑坡灾害点及隐患点进行叠加分析,然后进行统计分析,得到最终的试验结果。

在单独 I 模型的评价结果中,极低易发区和低易发区分别占整个研究区总面积的 32.69%,28.67%,滑坡灾害点及隐患点占比分别为 4.35%,23.60%,滑坡灾害点及隐患点密度分别 0.855,5.291 个/100 km<sup>2</sup>;而高易发区仅占整个研究区总面积的 16.59%,滑坡灾害点及隐患点占比却达到了 47.20%,滑坡灾害点及隐患点密度也高达 18.288 个/100 km<sup>2</sup>,具体如图 4 和表 3 所示。I 模型的频率比随着滑坡易发性等级的提升逐渐提升,且高易发性等级的频率比占总频率比的 57.73%,表明 I 模型能精确的对贵池区进行滑坡易发性评价。

在单独 LR 模型的评价结果中,极低易发区和低易发区分别占整个研究区总面积的 30.89%,27.53%,滑坡灾害点及隐患点占比分别为 6.83%,19.88%,滑坡灾害点及隐患点密度分别 1.421,4.639 个/100 km<sup>2</sup>;而高易发区仅占整个研究区总面积的 18.65%,滑坡灾害点及隐患点占比却达到了 47.83%,滑坡灾害点及隐患点密度也高达 16.478 个/100 km<sup>2</sup>(如图 4 和表 3 所示)。LR 模型的频率比随着滑坡易发性等级的提升逐渐提升,且高易发性等级的频率比占总频率比的 55.53%,表明 LR 模型能精确的对贵池区进行滑坡易发性评价。

在 I-LR 模型的评价结果中,极低易发区和低易发区分别占整个研究区总面积的 36.76%,23.96%,滑坡灾害点及隐患点占比分别为 5.59%,17.39%,滑坡灾害点及隐患点密度分别 0.977,4.664 个/100 km<sup>2</sup>;而高易发区仅占整个研究区总面积的 15.95%,滑坡灾害点及隐患点占比却达到了 50.93%,滑坡灾害点及隐患点密度也高达 20.526 个/100 km<sup>2</sup>,具体如图 4 和表 3 所示。I-LR 模型的频率比随着滑坡易发性等级的提升逐渐提升,高易发性等级的频率比占总频率比的 61.53%,表明 I-LR 模型能精确的对贵池区进行滑坡易发性评价。

在 I-GBDT-LR 模型的评价结果中,极低易发区和低易发区分别占整个研究区总面积的 35.55%,30.80%,滑坡灾害点及隐患点比例分别为 3.73%,19.25%,滑坡灾害点及隐患点密度分别 0.674,4.017 个/100 km<sup>2</sup>;而高易发区仅占整个研究区总面积的 11.36%,滑坡灾害点及隐患点比例却达到了 49.07%,滑坡灾害点及隐患点密度也高达 27.748 个/100 km<sup>2</sup>(如图 4 和表 3 所示)。I-GBDT-LR 耦合模型的频率比随着滑坡易发性等级的提升逐渐提升,高易发性等级的频率比占总频率比的 68.52%,表明 I-GBDT-LR 模型能精确的对贵池区进行滑坡易发性评价。



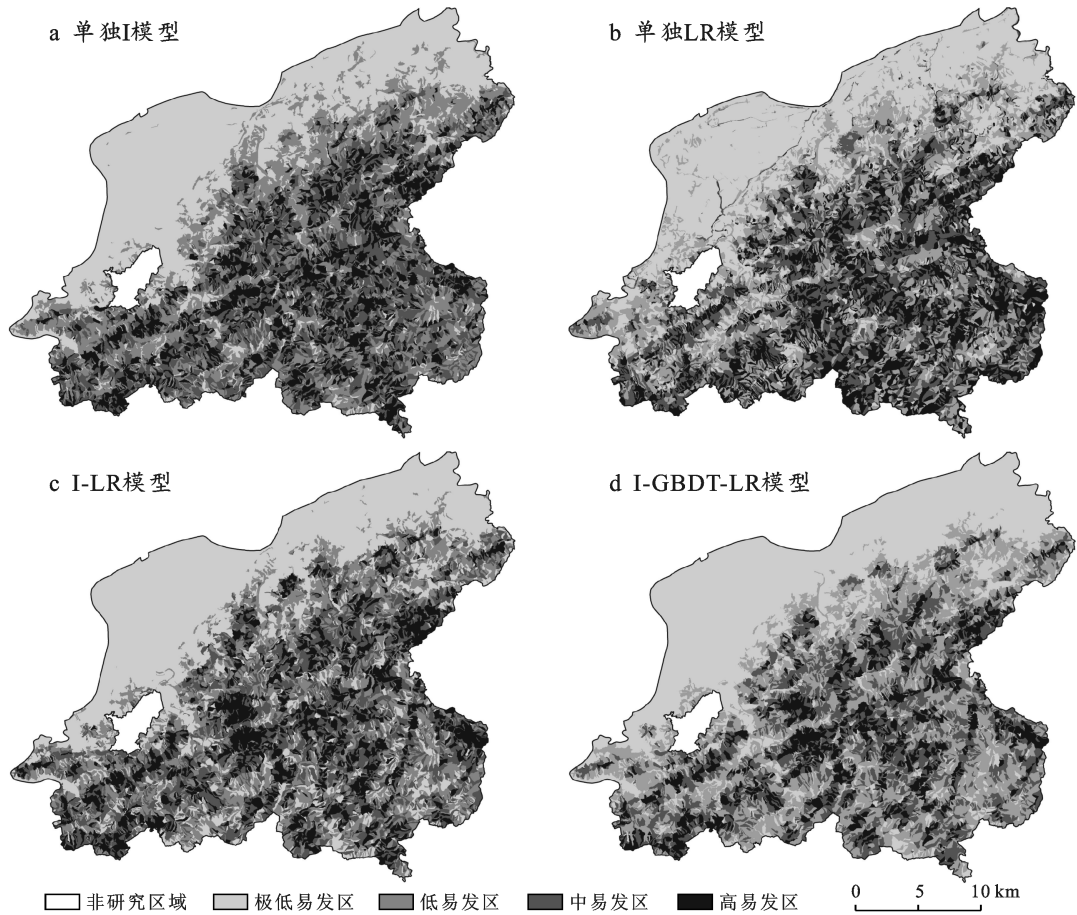


图 4 贵池区滑坡易发性区划

Fig.4 Landslide susceptibility zoning in Guichi District

表 3 频率比和灾害点及隐患点密度检验结果

Table 3 Frequency ratio and density test results of hazard points and hidden danger points

模型名称	易发性等级	面积/ km <sup>2</sup>	面积 比例(B)/%	滑坡点及 隐患点个数	滑坡点及隐患 点比例(A)/%	频率比 (A/B)	频率比 比例(C)/%	灾害点及隐患点密度/ (个·10 <sup>-2</sup> km <sup>-2</sup> )
I	极低易发区	818.992	32.69	7	4.35	0.133	2.70	0.855
	低易发区	718.133	28.67	38	23.60	0.823	16.70	5.291
	中易发区	552.427	22.05	40	24.84	1.127	22.87	7.241
	高易发区	415.575	16.59	76	47.20	2.845	57.73	18.288
LR	极低易发区	773.882	30.89	11	6.83	0.221	4.78	1.421
	低易发区	689.740	27.53	32	19.88	0.722	15.63	4.639
	中易发区	574.223	22.92	41	25.46	1.111	24.05	7.140
	高易发区	467.281	18.65	77	47.83	2.565	55.53	16.478
I-LR	极低易发区	920.826	36.76	9	5.59	0.152	2.93	0.977
	低易发区	600.281	23.96	28	17.39	0.726	13.99	4.664
	中易发区	584.531	23.33	42	26.09	1.118	21.55	7.185
	高易发区	399.489	15.95	82	50.93	3.193	61.53	20.526
I-GBDT-LR	极低易发区	890.684	35.55	6	3.73	0.105	1.67	0.674
	低易发区	771.651	30.80	31	19.25	0.625	9.91	4.017
	中易发区	558.089	22.28	45	27.95	1.255	19.90	8.063
	高易发区	284.702	11.36	79	49.07	4.320	68.52	27.748

### 3.4 评价模型的对比与分析

根据表 3 和 ROC 精度曲线,对比分析 4 种模型可知,在 4 种模型的频率比中,I-GBDT-LR 模型的高

易发频率比达 4.320,是所有模型中最高的,并且该模型的高易发区频率比占总频率比的 68.52%,也是 4 种模型中最高的。此外,4 种滑坡灾害点密度均是随



着易发性等级的提升而增大,并且高易发区的滑坡灾害点及隐患点密度显著高于其他易发性等级,I-GBDT-LR 模型相比于 I,LR 和 I-LR 模型,其高易发区的滑坡灾害点及隐患点密度分别提升约 9,11,7 个/100 km<sup>2</sup>。从 I-GBDT-LR 的 ROC 精度分别比 I,LR 和 I-LR 模型提升约 10%,9%,5%。

综上所述,从检验指标的各方面分析,虽然 I,LR 和 I-LR 模型,都能合理精确的对贵池区进行滑坡易发性评价,但是相比之下,本文所提出的 I-GBDT-LR 的耦合模型所得贵池区易发性评价结果是 4 种评价模型中最客观合理精确的。

## 4 结论

本文以安徽省池州市贵池区为研究区域,基于斜坡评价单元,选取坡度、坡向、地形曲率、剖面曲率、平面曲率、地形湿度指数(TWI)、归一化植被指数(NDVI)、地形起伏度、距断裂距离和距水系距离 10 个影响因子作为滑坡易发性评价指标,采用 I 模型、LR 模型、I-LR 模型和 I-GBDT-LR 模型,并基于 GIS 技术对贵池区进行滑坡易发性评价。

(1) 试验所采取的 4 种模型,从频率比、滑坡灾害点及隐患点密度和 ROC 精度曲线 3 个方面进行合理性和精确性检验,最终得到试验结果表明:4 种模型均能合理精确对贵池区进行滑坡易发性评价,所得滑坡易发性评价结果可以为政府部门防灾减灾工作提供有力的科学依据。

(2) 综合分析各评价模型可以得出,2 种单一模型的各方面检验指标均小于另 2 种耦合模型,在 4 种模型中 I-GBDT-LR 耦合模型的试验结果从频率比、滑坡灾害点及隐患点密度和 ROC 曲线 3 个方面进行检验,所得结果都是最好的,表明 GBDT 算法可以深入挖掘评价因子的特征信息,将其中的非线性特征转换成线性特征,再输入到 LR 模型中从而最终提升试验结果的精度,进一步提升滑坡易发性区划的准确度,与本文引言部分的理论分析相吻合。

由于试验条件的限制,试验中所用试验数据精度不一致、数据精度低、未收集到贵池区的地质数据、滑坡样本点不足等问题,可能都会在一定程度上影响最终的试验结果,在后续的研究中将进一步聚焦这些问题。

### [ 参 考 文 献 ]

[1] 殷坤龙,朱良峰.滑坡灾害空间区划及 GIS 应用研究[J]. 地质前缘,2001,8(2):279-284.  
 [2] 朱庆,张曼迪,丁雨淋,等.环境因子空间特征约束的区域滑坡敏感性模糊逻辑分析方法[J].武汉大学学报(信息

科学版),2021,46(10):1431-1440.

- [3] 郭子正,殷坤龙,黄发明,等.基于滑坡分类和加权频率比模型的滑坡易发性评价[J].岩石力学与工程学报,2019,38(2):287-300.  
 [4] 赵建华,陈汉林,杨树锋.滑坡灾害危险性评价模型比较[J].自然灾害学报,2006,15(1):128-134.  
 [5] 罗路广,裴向军,崔圣华,等.九寨沟地震滑坡易发性评价因子组合选取研究[J].岩石力学与工程学报,2021,40(11):2306-2319.  
 [6] 孙小凡,张鹏,党超.基于 GIS 的城市滑坡灾害易发性评价:以湖北省宜昌市城区为例[J].水土保持通报,2018,38(6):304-309.  
 [7] 许英姿,卢玉南,李东阳,等.基于 GIS 和信息量模型的广西花岗岩分布区滑坡易发性评价[J].工程地质学报,2016,24(4):693-703.  
 [8] 杜国梁,杨志华,袁颖,等.基于逻辑回归—信息量的川藏交通廊道滑坡易发性评价[J].水文地质工程地质,2021,48(5):102-111.  
 [9] 申怀飞,董雨,杨梅,等.基于 AHP 与信息量法的甘肃省滑坡易发性评估[J].水土保持研究,2021,28(6):412-419.  
 [10] 崔志超,王俊豪,崔传峰,等.基于层次分析法和模糊数学相结合的甘肃东乡八丹沟泥石流易发性评价[J].中国地质灾害与防治学报,2020,31(1):44-50.  
 [11] 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.  
 [12] 刘福臻,王灵,肖东升.机器学习模型在滑坡易发性评价中的应用[J].中国地质灾害与防治学报,2021,32(6):98-106.  
 [13] 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.  
 [14] 王世宝,庄建琦,郑佳,等.基于深度学习的 CZ 铁路康定—理塘段滑坡易发性评价[J].工程地质学报,2022,30(3):908-919.  
 [15] Polat A. An innovative, fast method for landslide susceptibility mapping using GIS-based LSAT toolbox [J]. Environmental Earth Sciences, 2021,80(6):217.  
 [16] Hung L Q, Van N T H, Duc D M, et al. Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: A case study in the upper Lo River Catchment (Vietnam) [J]. Landslides, 2016,13(5):1285-1301.  
 [17] Fang Zhice, Wang Yi, Duan Hexiang, et al. Comparison of general kernel, multiple kernel, infinite ensemble and semi-supervised support vector machines for landslide susceptibility prediction [J]. Stochastic Environmental Research and Risk Assessment, 2022, 36(10):3535-3556.