



DOI:10.12404/j.issn.1671-1815.2402758

引用格式:刘山东,李军,江兴元,等.基于数学统计与机器学习模型耦合的滑坡易发性评价方法优化[J].科学技术与工程,2025,25(5):1827-1839.

Liu Shandong, Li Jun, Jiang Xingyuan, et al. Optimization of landslide susceptibility assessment method coupling mathematical statistics and machine learning models[J]. Science Technology and Engineering, 2025, 25(5): 1827-1839.

基于数学统计与机器学习模型耦合的 滑坡易发性评价方法优化

刘山东¹, 李军², 江兴元^{1,3*}, 杨义^{1,3}, 赵荣乾³

(1. 喀斯特地质资源与环境教育部重点实验室(贵州大学), 贵阳 550025; 2. 贵州省地质矿产勘查开发局 114 地质大队, 遵义 563000; 3. 贵州大学资源与环境工程学院, 贵阳 550025)

摘要 滑坡地质灾害易发性评价是防灾减灾的一种重要手段,易发性评价模型的选取和优化至关重要。以思南县为研究区,选取高程、坡度、曲率、地层、土地利用、年平均降雨量等 16 个评价因子,采用频率比(frequency ratio,FR)模型与支持向量机(support vector machine,SVM)模型和随机森林(random forest,RF)模型相耦合,引入网格搜索方法来获取 SVM 模型、RF 模型及其耦合模型最优参数组合并用于模型训练,最终构建 SVM、RF、FR-SVM 及 FR-RF 模型对整个研究区进行滑坡易发性预测,并进行了受试者操作特征(receiver operating characteristics,ROC)曲线验证。结果表明:与单一机器学习模型相比,耦合机器学习有更多的滑坡灾害样本落于高易发区和极高易发区,有更高的准确率。单一模型中,RF 模型有较多的滑坡灾害样本落于高易发区和极高易发区,耦合模型中,FR-RF 模型有较多的滑坡灾害样本落于高易发区和极高易发区,且 FR 模型和 FR-RF 模型中没有滑坡灾害样本落在极低易发区,表明无论是单一模型还是耦合模型,RF 模型的性能优于 SVM 模型。4 种模型的 ROC 预测曲线的曲线下面积(area under the curve,AUC)分别为 0.831 6、0.843 9、0.864 4、0.910 4,说明 FR 模型与 RF 模型结合的耦合模型有更高的准确率,该模型更适用于思南县的滑坡易发性评价研究,评价结果可为当地滑坡地质灾害的防灾减灾提供一定的参考。

关键词 滑坡易发性评价;频率比模型;机器学习模型;耦合模型;ROC 曲线;思南县

中图法分类号 P642.2; 文献标志码 A

Optimization of Landslide Susceptibility Assessment Method Coupling Mathematical Statistics and Machine Learning Models

LIU Shan-dong¹, LI Jun², JIANG Xing-yuan^{1,3*}, YANG Yi^{1,3}, ZHAO Rong-qian³

(1. Key Laboratory of Karst Georesources and Environment (Guizhou University), Ministry of Education, Guiyang 550025, China;

2. 114 Geological Brigade, Guizhou Geological and Mining Bureau, Zunyi 563000, China;

3. College of Resources and Environmental Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] Landslide geological hazard susceptibility assessment is an important means of hazard prevention and reduction. The selection and optimization of susceptibility assessment model is very important. Sinan County was selected as the study area, and 16 assessment factors such as elevation, slope, curvature, lithology, land use, and average annual precipitation were selected. Frequency ratio (FR) model was coupled with support vector machine (SVM) model and random forest (RF) model. Grid search method was introduced to obtain the optimal parameter combination of SVM model, RF model and their coupling model for model training. Finally, SVM, RF, FR-SVM and FR-RF models were constructed to predict landslide susceptibility in the whole study area, and receiver operating characteristics (ROC) curve was performed verification. The results show that compared with the single machine learning model, the coupled machine learning model has more landslide hazard samples fall in the high zone and the very high zone, and has higher accuracy. In the single model, more landslide hazard samples in the RF model fall in the high zone and the extremely high zone. In the coupled model, more landslide hazard samples in the FR-RF model fall in the high zone and the very high zone, and no hazard samples points in the FR model and the FR-RF model fall in the very low zone, indicating that no matter the single model or the coupled model,

收稿日期:2024-04-16 修订日期:2024-11-20

基金项目:国家自然科学基金(42007271);贵州省科技支撑计划项目(黔科合支撑[2023]一般119)

第一作者:刘山东(2000—),男,汉族,贵州毕节人,硕士研究生。研究方向:InSAR、地质灾害易发性评价。E-mail:liushandonggiser@163.com。

*通信作者:江兴元(1986—),男,汉族,山东菏泽人,博士,副教授。研究方向:地质工程、岩土工程。E-mail:xyjiang3@gzu.edu.cn。

The performance of RF model is better than that of SVM model. The AUC values of ROC prediction curves of the four models are 0.831 6, 0.843 9, 0.864 4 and 0.910 4, indicating that the coupling model combined with FR model and RF model has a higher accuracy, and this model is more suitable for the assessment of landslide susceptibility in Sinan County. The assessment results can provide some reference for hazard prevention and reduction of local landslide geological hazards.

[**Keywords**] landslide susceptibility assessment; frequency ratio model; machine learning models; coupled model; receiver operating characteristic curve; Sinan County

滑坡具有破坏性极强、发生频繁、规模大、分布广以及隐蔽性强等特点,滑坡灾害发生造成严重的人员伤亡和财产损失难以估量,因此对滑坡灾害进行易发性评价是预防滑坡灾害发生的必要手段^[1-2]。滑坡易发性评价是对某一地区内潜在滑坡的发生可能性进行评估和预测发生滑坡的概率的过程^[3]。通过综合考虑地形、地质、水文、土地覆盖以及人类活动等因素,根据已发生滑坡的环境特征,评估出该地区的地质环境和自然条件对滑坡形成的影响,从而确定该地区的滑坡易发性等级,并为防范滑坡发生灾害提供一定的科学依据^[4-5]。

当前滑坡易发性评价模型主要有信息量^[6-7]、频率比^[8]、确定性系数^[9]、层次分析^[10-11]、逻辑回归^[12]、支持向量机^[13]、随机森林^[14]、人工神经网络^[15]以及多种评价方法的耦合模型等。学者对此进行了大量的实验和研究。Huang等^[16]使用频率比和支持向量机模型在中国东南丘陵南田地区开展滑坡易发性制图研究,根据多种影响因子生成滑坡易发性区划图,并且对两种模型的精度和效率进行评估和比较。Fang等^[17]使用支持向量机及其流行的变体(即多核学习、无限集成支持向量机和半监督支持向量机)来预测滑坡的易发性概率,并比较其预测性能。Shou等^[18]使用逻辑回归和支持向量机模型对中国台湾省根据降雨的差异进行极端降雨预测以及对可能引起的滑坡进行易发性评估,并且比较和讨论了不同模型的局限性。陈芯宇等^[19]基于确定性系数与支持向量机模型采取灰狼优化算法得到最优参数建立确定性系数-支持向量机模型对略阳县进行易发性预测,通过受试者特征工作曲线下的面积比较可知,确定性系数-支持向量机模型比单一的支持向量机模型有更高的准确率。Luo等^[20]使用人工神经网络、支持向量机和信息量模型对矿山附近因生态脆弱性可能导致的滑坡进行了易发性评价,将估算的滑坡易发性区划图与地面真实情况进行比较,高易发区往往位于具有多个断层分布和陡峭斜坡的中部地区,与真实情况比较符合。

现应用频率比模型和机器学习耦合的方法来评估滑坡易发性,多模型耦合的主要优点是可以综合考虑不同单一模型的优势,从而提高预测和评价的准确性和可靠性,同时多模型耦合还可以增加对

影响滑坡易发性的多种因素的综合分析,更全面地评价滑坡的易发程度^[21-22]。选取高程、坡向、坡度、曲率、平面曲率、剖面曲率、地形湿度指数、距断层距离、地层、土地利用、归一化差异植被指数、距道路距离、距水系距离、年平均降雨量等16个影响因子作为易发性评价因子,使用频率比模型和支持向量机模型以及随机森林模型相结合的耦合模型得到研究区的易发性评价结果,生成的易发性评价分区图可为当地防灾减灾提供参考。

1 滑坡易发性评价模型

1.1 频率比模型

频率比(frequency ratio, FR)是一种简单实现且结果准确的双变量统计方法,该方法是基于观测到的滑坡分布与各滑坡相关因子之间的关系,揭示了研究区滑坡位置与致灾因子之间的相关性^[8,16]。对于一些连续性的易发性评价因子(如高程、坡度等),结合前人的经验以及野外调查进行重分类分级,对于固有分级的分类因子(如土地利用、地层岩性等)按照原有的分类,最后使用频率比模型中的式(1)计算易发性评价因子每个类中发生滑坡的单元格比例,并将计算得到每个类的频率比值赋给每个因子分类。

$$FR_{ij} = \frac{N_{ij}/N}{S_{ij}/S} \quad (1)$$

式(1)中: i 为易发性评价因子; j 为易发性评价因子下的分级; FR_{ij} 为第 i 个易发性评价因子中第 j 个分级的频率比值; N_{ij} 为第 i 个易发性评价因子,第 j 个分级的滑坡数; N 为研究区的滑坡总数; S_{ij} 为第 i 个易发性评价因子,第 j 个分级的栅格数; S 为研究区的总栅格数。

1.2 支持向量机模型

支持向量机(support vector machine, SVM)模型是一种随着人工智能的发展和广泛应用而兴起的一种机器学习算法,具有使用样本少、非线性和高维数等特点^[13]。基本思想是将高维空间中的数据点划分到不同的类别中,以此来建立预测模型,并在特征空间中找到一个最优的超平面,使得不同类别的数据点被分到不同的一侧,并且距离超平面最近的数据点与超平面的距离最大化^[14],原理如图1所示。

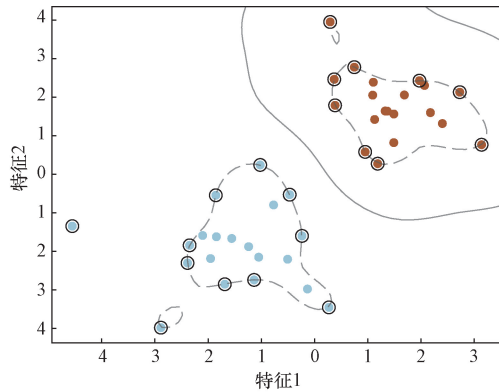


图1 支持向量机模型原理

Fig. 1 The principle of support vector machine model

随着多学科的融合发展,支持向量机逐渐被应用到地质灾害的易发性评价中。支持向量机模型可以用于分类问题,模型基于训练样本的特征构建超平面,使得不同类别的样本能够被最大化地分离,对不同的滑坡样本和非滑坡样本进行区分和预测。通过建立训练样本集合和测试样本集合,对模型进行训练和测试,可以得到一个较为准确的滑坡易发性预测模型。

1.3 随机森林模型

随机森林(random forest, RF)模型是一种基于决策树的集成学习方法,该方法利用从大量样本数据集中生成的多分类回归树来确定每个实验样本属于哪一类,每棵决策树代表一个分类器基于决策树构建多个决策树模型,然后使用投票或平均法来对每个子样本进行建模,最终将所有模型汇总,并通过投票过程确定最佳模型^[14],原理如图2所示。

随机森林模型在处理高维数据和非线性问题方面表现出色。它们在处理大规模数据集时也表现出显著的效果,还可以评估特征的重要性,有助于优化特征选择过程,更好地理解滑坡敏感性与各种因素之间的复杂关系。当前随机森林模型已经成为最常用且高度准确的滑坡易发性评价机器学习方法之一。

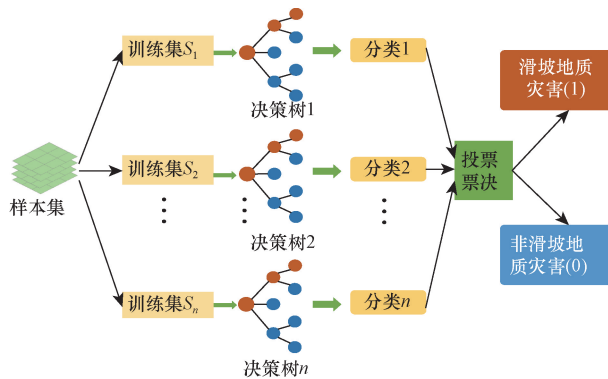


图2 随机森林模型原理

Fig. 2 The principle of random forest model

1.4 耦合模型

将FR模型分别与SVM模型和RF进行耦合,主要原理是将FR模型计算得到的易发性评价因子不同分级每个栅格单元的频率比值作为SVM模型和RF模型易发性评价因子的输入值,最终构建频率比-支持向量机模型(FR-SVM)和频率比-随机森林模型(FR-RF)。将FR模型分别与SVM模型和RF模型进行耦合,将FR模型计算得到的不同评价因子所计算得到的信息量和确定性系数作为支持向量机模型和随机森林模型的输入评价因子,这样可以解决评价模型中不同评价因子之间数据不整合的问题,从而提高模型预测准确性和可靠性,降低易发性评价模型中的人为主观性的影响。原理如图3所示。

2 研究区概况

思南县位于贵州省东北部,坐标为(107°52'22"E ~ 108°27'24"E, 27°31'42"N ~ 28°9'24"N),地处武陵山腹地,是乌江流域的中心地带,总面积约2 230.5 km²。区域内地势起伏,山峰和山谷交错,河流众多,山丘层叠,如图4所示。受构造、岩性和乌江水系水文特征的影响,该地区呈现出丰富的地貌形态,包括高耸的山峰、缓坡的山谷和蜿蜒的河谷^[23]。由于境内山高坡陡、地形切割大、地质构造发育、地层岩性及其组合多样、地质环境条件复杂,以及近年来气候多变、降雨量大等条件,导致该县滑坡地质灾害频繁发生,是贵州省地质灾害多发县之一。

3 评价因子的建立与分析

3.1 评价因子选取

滑坡灾害数据来源于思南县历年的地质灾害野外调查资料,采用了遥感影像解译、历史记录和实地调查相结合的方法,生成详细的滑坡灾害分布图(图3)。结果显示研究区域内存在200个滑坡地质灾害,滑坡总面积为5.24 km²,最大滑坡面积为248 525.9 m²,最小滑坡面积为172.57 m²;高程数据来源于地理空间数据云的ASTER GDEM (<https://www.gscloud.cn/>),空间分辨率为30 m,同时还用于易发性评价因子坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度、地表粗糙度以及地形湿度指数的提取;研究所用的地质图数据来源于地球科学数据出版系统(<http://dce.cgs.gov.cn/>),将地质图矢量化用于地层和断层的提取;空间分辨率为30 m的Landsat 8 OLI_TIRS卫星遥感影像,用于研究区归一化差异植被指数(normalized difference vegetable index, NDVI)提取;土地利用类型数据来源

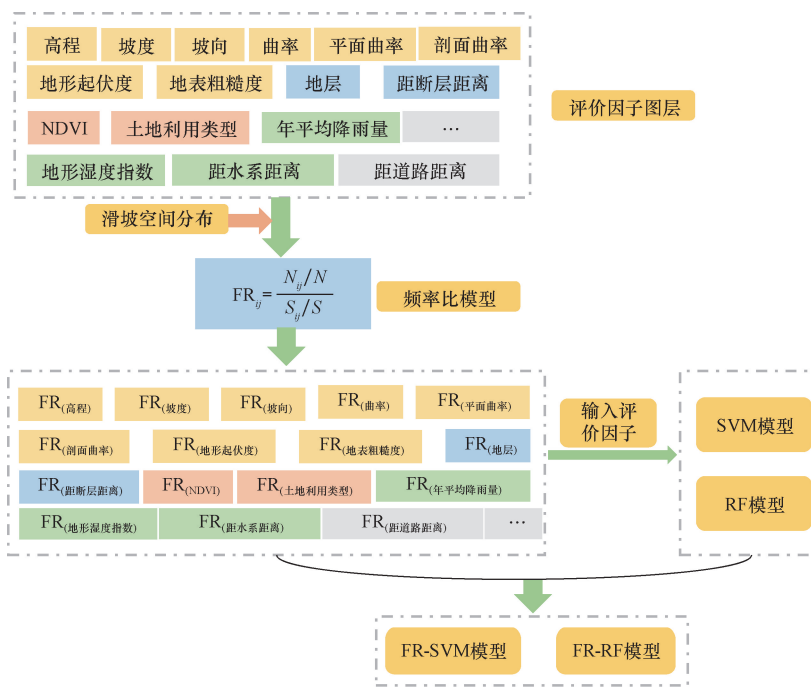
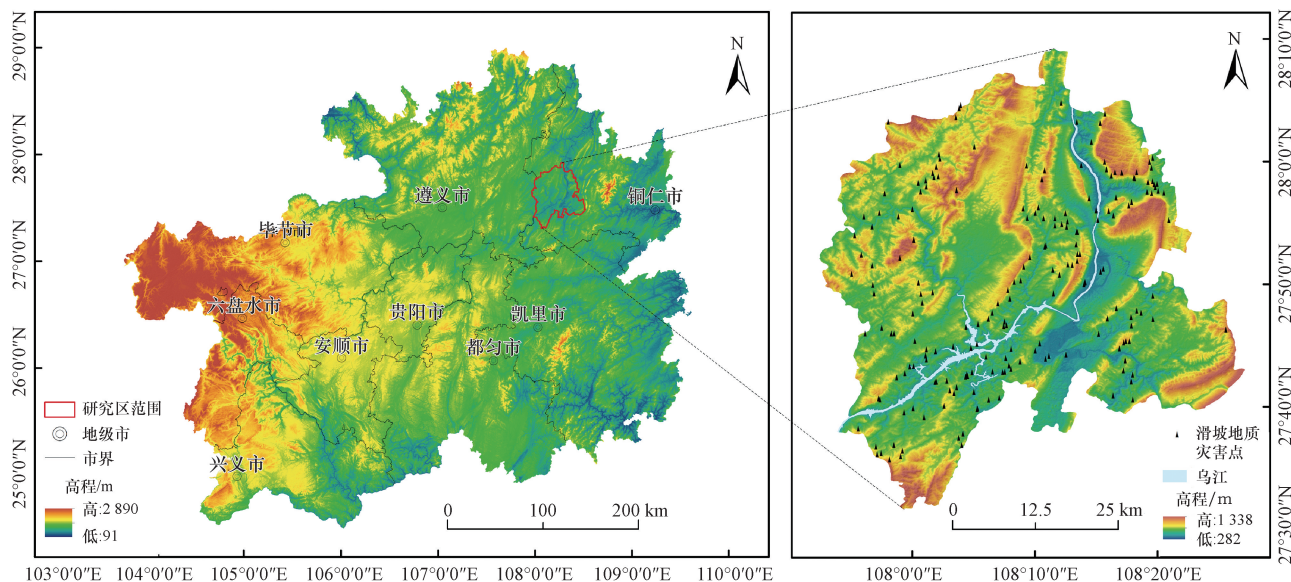


图3 耦合模型原理

Fig. 3 The principle of couple model



审图号:GS(2020)4619号

图4 研究区地理位置及滑坡灾害分布图

Fig. 4 Geographical localization and landslide hazard distribution map of the study area

于全球地理信息产品 Globeland30 (<http://www.globallandcover.com/>), 数据年份为2020年; 水系和道路数据来源于全国地理信息资源目录服务系统 1:100万全国基础地理数据库 (<https://www.webmap.cn/>); 降雨数据来源于美国国家海洋和大气管理局(NOAA), 数据年份分别为2020年、2021年和2022年气象站点日统计数据, 将每个站点计算平均值得到年均降雨量, 再对其进行克里金插值得到研

究区年均降雨量等值线图。

根据滑坡地质灾害发生与地质环境之间的关联性, 主要从高程、坡度、坡向、降雨等16个方面选取致灾因素作为滑坡易发性评价因子。为了便于统计和计算, 选取规则的栅格单元作为评价单元, 根据计算栅格单元大小的经验公式^[24]。

$$G_s = 7.49 + 0.006S - 2.0 \times 10^{-9}S^2 + 2.9 \times 10^{-15}S^3 \quad (2)$$

式(2)中: G_s 为适宜网格大小; S 为原始等高线数据精度的分母。

通过计算最佳网格单元为 32.8 m,因此将栅格单元的尺寸划分为 30 m × 30 m,根据易发性评价因子采用频率比模型与机器学习模型耦合的评价模型对研究区的滑坡地质灾害进行滑坡易发性评价。

滑坡灾害易发性评价的精度主要受评价因子的影响,因此合理地选取评价因子十分重要^[25]。在保证环境因子客观、准确的前提下,结合野外现场地质调查。选择了高程、坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度、地表粗糙度、地层岩性、距断层距离、距水系距离、距道路距离、地形湿度指数、年均降雨量、NDVI 和土地利用类型等 16 个评价因子。将所有评价因子分为 5 个类别。

- (1) 地形地貌因子(高程、坡度和坡向等^[26])。
- (2) 基础地质因子(地层和距断层距离)。
- (3) 土地覆盖因子(土地利用类型和 NDVI)。
- (4) 水文环境因子(距水系距离、年平均降雨量和地形湿度指数)。
- (5) 人类活动因子(距道路距离)。

其中地层和土地利用为离散型评价因子,其余 14 个评价因子属于连续型数据。离散型评价因子可以根据其原始属性组进行分类,而连续型评价因子需根据研究区域的实际情况和以往研究经验进行分级,如图 5 所示。

3.2 相关性分析

在滑坡易发性评价过程中,为保证模型的准确性,应剔除相关性较高的评价因子。通过计算评价因子之间的皮尔逊相关系数(pearson correlation coefficient, PCC)来衡量因子之间的相关性^[22],计算公式为

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

式(3)中: r 为所计算的皮尔逊相关系数; X_i 和 Y_i 分别是两个评价因子的第 i 个观测值; \bar{X} 和 \bar{Y} 分别为两个条件因子的平均值; n 为评价因子的数量。

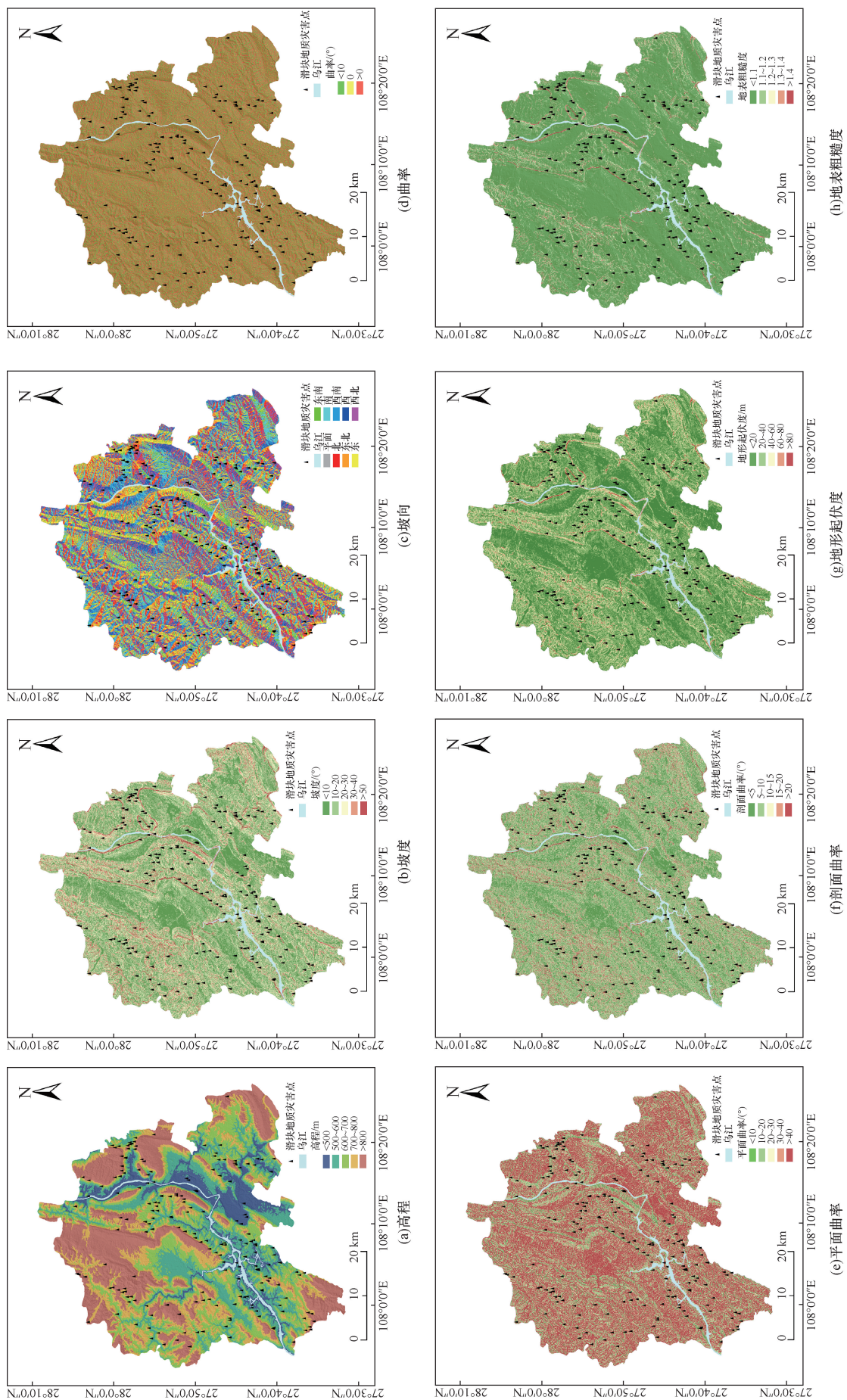
如果两个评价因子之间的皮尔逊相关系数绝对值小于 0.5,表明这两个评价因子之间的相关性较低,可以保留这两个评价因子。反之,如果两个评价因子之间的相关系数绝对值大于 0.5,说明这两个评价因子之间的相关性较高,需要舍弃其中一个评价因子。同时通过计算所有评价因子之间的信息增益比(information gain ratio, IGR)来评估评价因子的重要性,信息增益比方法是常用的特征选择和数据挖掘技术,用于评估不同条件因素对滑坡发生可能性的影响

程度。该方法基于信息熵的概念,旨在找到在预测滑坡敏感性时具有最强区分能力的因素,以提高预测的准确性^[27]。对于相关性较高的两个评价因子,可以根据计算得到的信息增益比值,舍弃掉重要性较低的评价因子。根据皮尔逊相关系数计算结果[图 6(a)]和信息增益比值计算结果[图 6(b)]可知,坡度、地形起伏度和地表粗糙度这 3 个评价因子之间存在较高的相关性,应舍弃地形起伏度和地表粗糙度这两个评价因子,保留 14 个易发性评价因子。

4 思南县滑坡易发性评价

4.1 基于机器学习模型的滑坡易发性评价结果

对思南县 200 个滑坡灾害进行半径为 1 km 的缓冲区分析,在缓冲区以外随机选择 200 个非滑坡地质灾害样本与滑坡地质灾害样本组成样本集,将样本集与 14 个易发性评价因子进行栅格提取得到样本集的评价因子空间数据,并将样本集分为 70% 的训练集和 30% 的测试集,在 Python 编程语言中进行支持向量机和随机森林模型的训练和验证。为了保证易发性评价模型的精度,两种模型均采用网格搜索的方法获取最优参数组合,支持向量机模型中核函数默认为径向基函数(radical basis function, RBF),设置的参数为惩罚因子 C 和 γ ,设置 C 取值范围为 0.01、0.02、0.05、0.08、0.1、0.2、0.5、0.8、1、2、5、8、10、20、32, γ 的取值范围为 0.01、0.02、0.05、0.08、0.1、0.2、0.5、0.8、1、2。通过受试者工作曲线(ROC)进行不同参数组合模型的精度评价,使用 ROC 曲线下面积(AUC)来衡量,该方法是用于评估二分类模型性能的一种常用工具^[28]。当 AUC 大于 0.5 时,AUC 越接近 1,精度越高,当 C 为 2, γ 为 0.01 时,AUC 最高,支持向量机模型效果最佳[图 7(a)]。随机森林模型中设置的参数为决策树数量($n_estimators$)、决策树最大深度(max_depth)、叶节点最少样本数($min_samples_leaf$)以及内部节点分裂所需最小样本数($min_samples_split$),设置决策树数量范围为 100、200、300、400、500、600、700、800、900、1 000、1 100、1 200、1 300、1 400、1 500、1 600、1 700、1 800、1 900、2 000,决策树最大深度范围为 10、20、30,叶节点最少样本数范围为 1、2、4,内部节点分裂所需最小样本数范围为 2、5、10。当决策树数量为 300,决策树深度为 10,叶节点最少样本数为 2,叶节点最少样本数为 2 时,AUC 最高,随机森林模型效果最佳[图 7(b)]。通过训练好的机器学习模型,计算得到思南县栅格单元的易发性指数,最后利用等间隔法将计算得到的易发性指数进行重分类,将研究区分为极高易发区、高易发区、中易发区、低易发区以及极低易发区[图 8(a)、图 8(b)]。



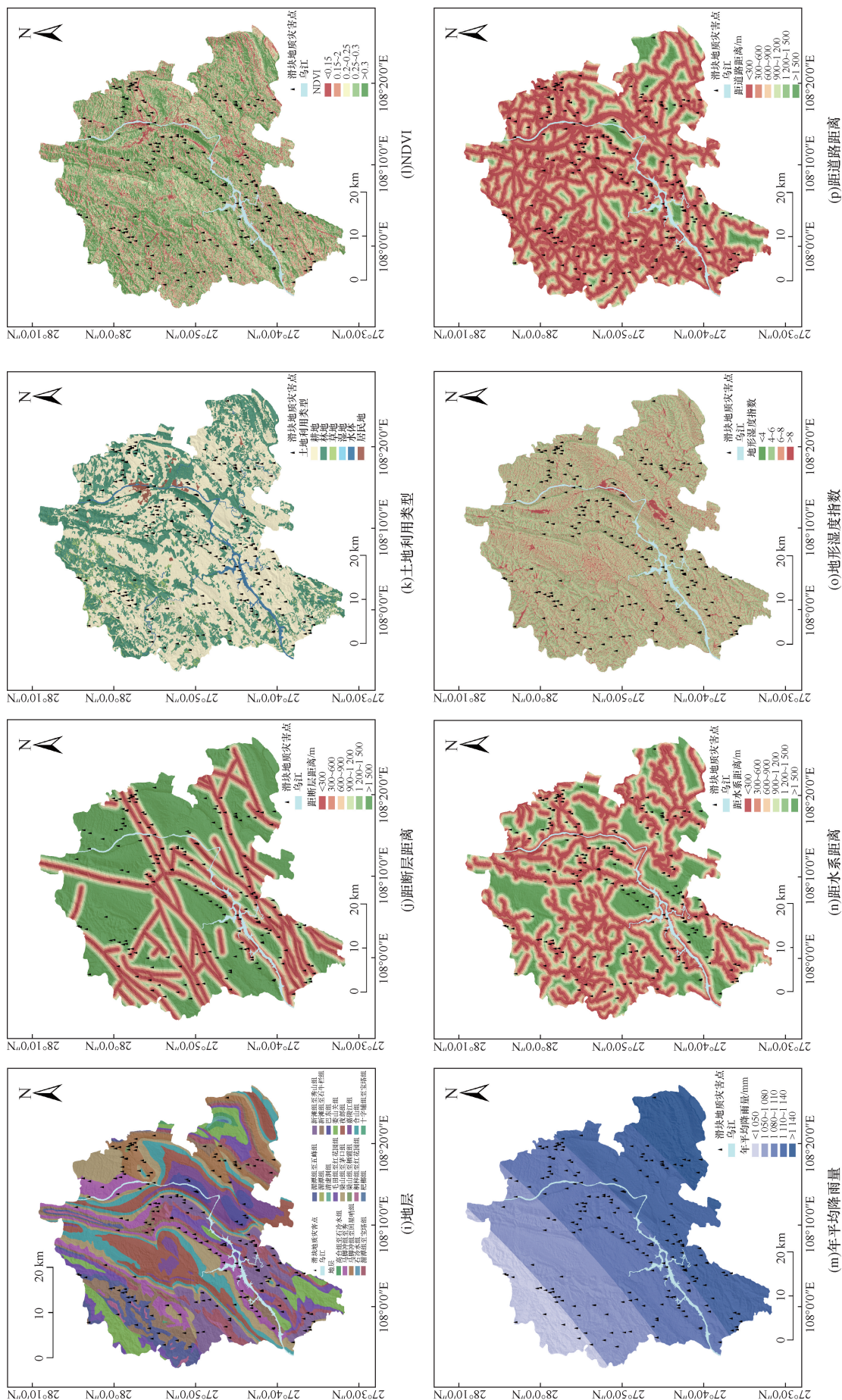


图 5 评价因子图 Fig. 5 Landslide susceptibility assessment factor map

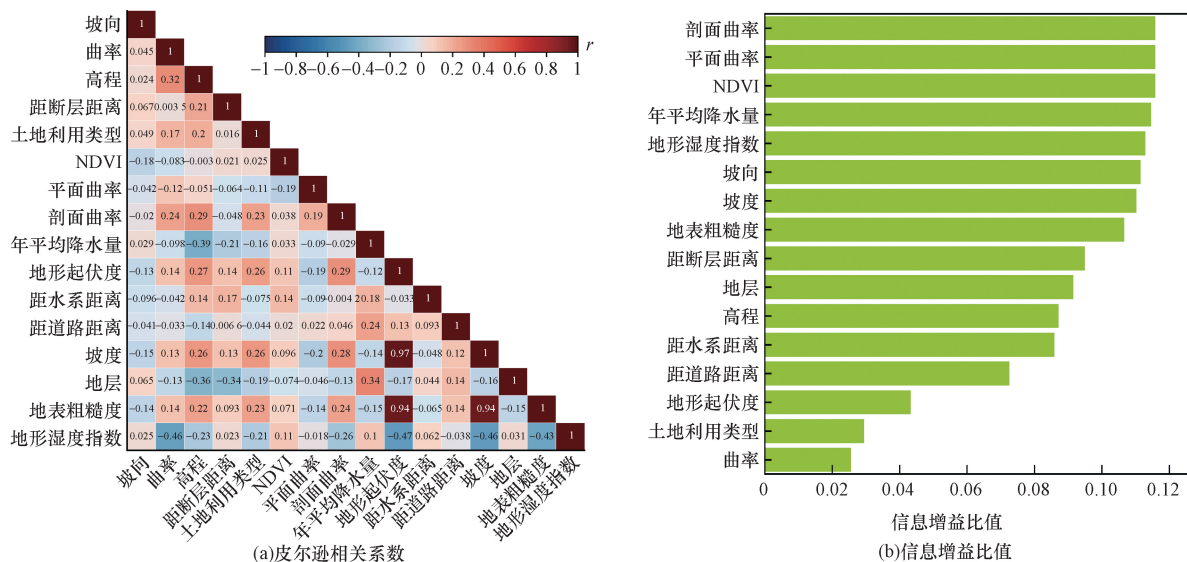


图6 各评价因子相关系数和信息增益比值
Fig. 6 Correlation coefficients and information gain ratios of each assessment factor

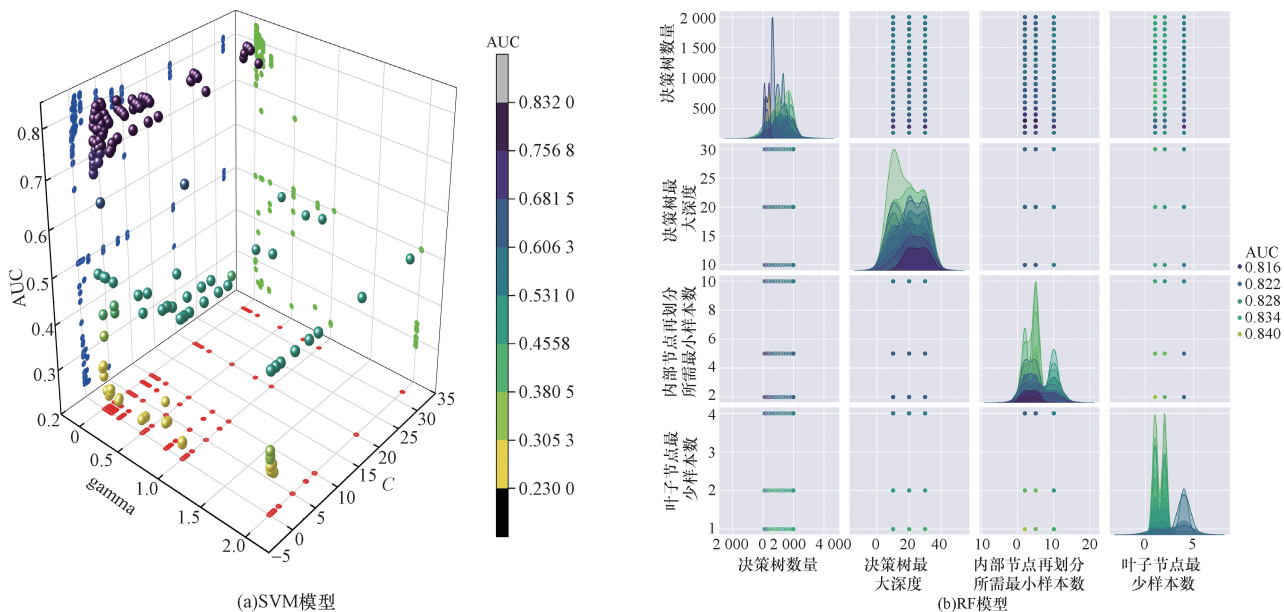


图7 SVM 和 RF 模型最优参数组合
Fig. 7 Optimal parameter combinations for SVM and RF models

4.2 基于频率比模型与机器学习模型耦合的滑坡易发性评价结果

使用式(1)计算不同评价因子在不同分级区间下的频率比值(表1),得到不同评价因子频率比值的空间分布图,并将其作为支持向量机模型和随机森林模型的输入评价因子,将样本集与不同评价因子频率比值的空间分布数据空间连接得到样本集的评价因子空间数据。同样按照 7:3 的比例分为训练集和测试集,在 Python 中均采用网格搜索的方法得到最佳的参数组合,参数设置与前面机器学习模

型中设置参数一致,当 C 为 2, gamma 为 0.01 时, AUC 最高,频率比-支持向量机(FR-SVM)模型效果最佳[图 9(a)]。当决策树数量为 1 700,决策树深度为 20,叶节点最少样本数为 5,叶节点最少样本数为 1 时,AUC 最高,频率比-随机森林(FR-RF)模型效果最佳[图 9(b)]。最终训练好的耦合机器学习模型,计算得到思南县栅格单元的易发性指数,最后利用等间隔法将计算得到的易发性指数进行重分类,将研究区分为极高易发区、高易发区、中易发区、低易发区以及极低易发区[图 10(a)、图 10(b)]。

表 1 不同评价因子的频率比值
Table 1 Frequency ratio values for different assessment factors

评价因子	分级	滑坡占比	栅格占比	频率比值	评价因子	分级	滑坡占比	栅格占比	频率比值
高程/m	<500	0.110 0	0.108 7	1.011 9	年均 降雨量/ mm	<1 050	0.065 0	0.096 0	0.677 0
	500~600	0.200 0	0.188 4	1.061 8		1 050~1 080	0.145 0	0.242 3	0.598 4
	600~700	0.275 0	0.233 7	1.177 0		1 080~1 110	0.465 0	0.313 4	1.483 7
	700~800	0.210 0	0.189 0	1.111 0		1 110~1 140	0.290 0	0.251 6	1.152 8
	>800	0.205 0	0.280 3	0.731 4		>1 140	0.035 0	0.096 7	0.362 0
坡度/(°)	<10	0.055 0	0.248 9	0.220 9	NDVI	<0.15	0.135 0	0.117 8	1.146 5
	10~20	0.480 0	0.412 0	1.165 1		0.15~0.2	0.160 0	0.135 4	1.181 6
	20~30	0.255 0	0.236 0	1.080 7		0.2~0.25	0.235 0	0.218 0	1.077 9
	30~40	0.110 0	0.079 8	1.377 6		0.25~0.3	0.225 0	0.253 0	0.889 2
	>40	0.100 0	0.023 3	4.292 3		>0.3	0.245 0	0.275 8	0.888 3
坡向	平面	0.000 0	0.003 6	0.000 0	土地利用 类型	耕地	0.770 0	0.578 9	1.330 2
	北	0.120 0	0.117 8	1.018 3		林地	0.190 0	0.353 2	0.537 9
	东北	0.120 0	0.115 3	1.040 7		草地	0.035 0	0.038 8	0.902 7
	东北	0.110 0	0.122 2	0.900 4		湿地	0.000 0	0.000 1	0.000 0
	东南	0.175 0	0.131 0	1.335 7		水体	0.005 0	0.021 5	0.233 0
	南	0.125 0	0.127 5	0.980 7		人造水体	0.000 0	0.007 6	0.000 0
	西南	0.115 0	0.120 3	0.955 8	地层	湄潭组至五峰组	0.095 0	0.048 0	1.979 5
	西	0.110 0	0.126 8	0.867 7		毛田组至红花园组	0.020 0	0.066 8	0.299 4
	西北	0.125 0	0.135 5	0.922 7		梁山组至茅口组	0.110 0	0.126 4	0.870 6
地形曲率	<0	0.465 0	0.482 0	0.964 8		马脚冲组至秀山组	0.165 0	0.088 9	1.857 0
	0	0.055 0	0.039 3	1.399 9		马脚冲组至回星哨组	0.135 0	0.055 0	2.452 5
	>0	0.480 0	0.478 7	1.002 6		合山组	0.005 0	0.091 3	0.054 8
剖面曲率	0~5	0.380 0	0.310 4	1.224 3		夜郎组	0.000 0	0.115 9	0.000 0
	5~10	0.320 0	0.378 9	0.844 6		娄山关组	0.005 0	0.073 4	0.068 1
	10~15	0.140 0	0.191 8	0.729 8		石冷水组	0.000 0	0.000 0	0.000 0
	15~20	0.080 0	0.078 3	1.022 1		新滩组至石牛栏组	0.070 0	0.080 5	0.869 8
	>20	0.080 0	0.040 6	1.968 2		湄潭组	0.000 0	0.000 4	0.000 0
平面曲率	<10	0.140 0	0.076 8	1.823 2		清虚洞组	0.000 0	0.001 8	0.000 0
	10~20	0.285 0	0.168 4	1.692 5	嘉陵江组	0.000 0	0.051 7	0.000 0	
	20~30	0.210 0	0.174 0	1.206 6	桐梓组至红花园组	0.020 0	0.010 3	1.949 6	
	30~40	0.165 0	0.143 3	1.151 1	高台组至石冷水组	0.000 0	0.004 7	0.000 0	
	>40	0.200 0	0.437 4	0.457 2	湄潭组至宝塔组	0.110 0	0.076 0	1.446 5	
地形湿度 指数	<4	0.080 0	0.028 6	2.793 6	梁山组至栖霞组	0.020 0	0.000 7	28.267 0	
	4~6	0.585 0	0.567 1	1.031 6	巴东组	0.050 0	0.023 0	2.172 6	
	6~8	0.245 0	0.262 8	0.932 3	新滩组至秀山组	0.195 0	0.084 4	2.310 0	
	>8	0.090 0	0.141 5	0.636 1	十字铺组至宝塔组	0.000 0	0.000 2	0.000 0	
距水系 距离/m	<300	0.170 0	0.258 6	0.657 3	榔榔组	0.000 0	0.000 6	0.000 0	
	300~600	0.260 0	0.208 4	1.247 5	距断层 距离/m	<300	0.260 0	0.160 9	1.616 1
	600~900	0.195 0	0.168 7	1.156 2		300~600	0.115 0	0.138 4	0.831 0
	900~1 200	0.200 0	0.126 4	1.581 8		600~900	0.130 0	0.114 3	1.137 2
	1 200~1 500	0.040 0	0.087 5	0.457 3		900~1 200	0.125 0	0.092 3	1.354 0
>1 500	0.135 0	0.150 4	0.897 6	1 200~1 500		0.080 0	0.077 0	1.039 3	
距道路 距离/m	<300	0.525 0	0.429 2	1.223 1	>1 500	0.290 0	0.417 1	0.695 2	
	300~600	0.205 0	0.264 2	0.775 9					
	600~900	0.125 0	0.160 1	0.780 7					
	900~1 200	0.100 0	0.082 3	1.215 6					
	1 200~1 500	0.030 0	0.036 1	0.830 9					
>1 500	0.015 0	0.028 1	0.534 7						

4.3 滑坡易发性评价模型对比分析

通过统计上述模型不同易发性分区的滑坡灾害数与栅格个数,并计算所占总滑坡数比例、所占总栅格数比例以及滑坡个数占比。如表 2 所示,

SVM、RF、FR-SVM、FR-RF 模型易发性分区的高易发区和极高易发区所包含的滑坡数占总滑坡数比例分别为 75.00%、87.00%、85.00%、91.00%,其中与频率比模型耦合的机器模型所包含的滑坡地质

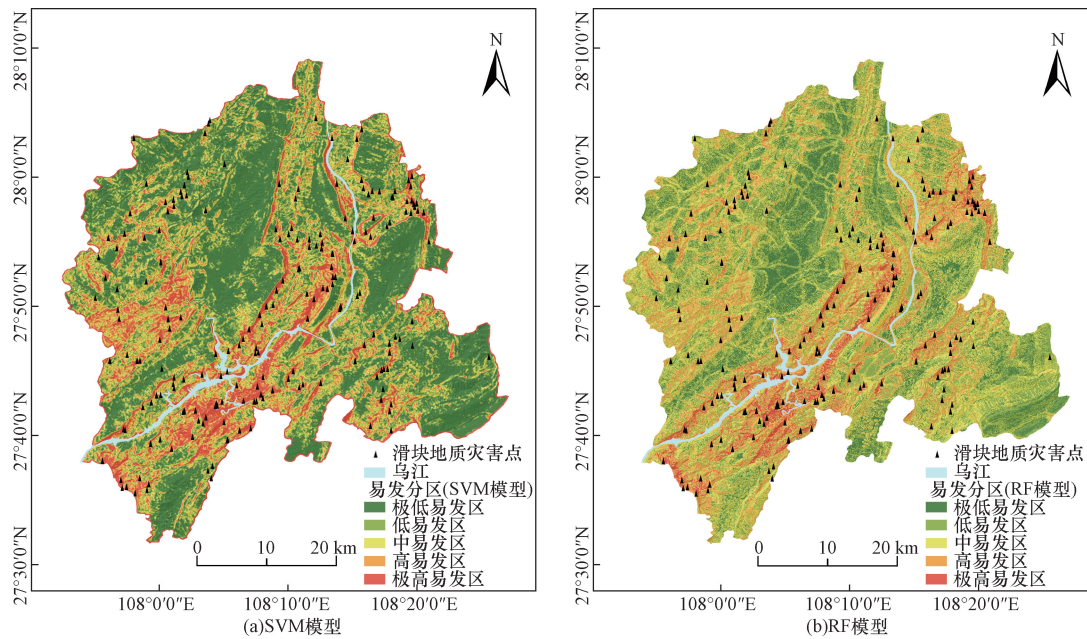


图8 基于SVM模型和RF模型的滑坡易发性分区图

Fig. 8 Landslide susceptibility zoning map based on SVM and RF models

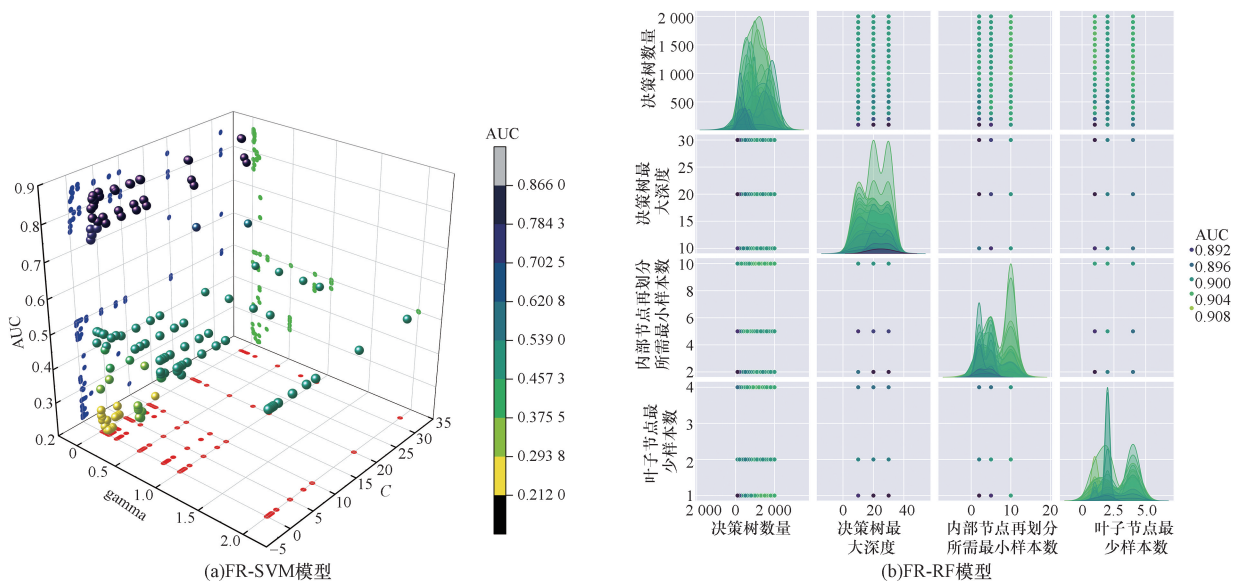


图9 FR-SVM模型和FR-RF模型最优参数组合

Fig. 9 Optimal parameter combinations for FR-SVM and FR-RF models

灾害占比均高于单一机器学习模型,表明耦合机器学习模型的性能是优于单一机器学习模型的。就单一机器学习模型而言,RF模型中没有滑坡灾害点落于极低易发区,且更多滑坡灾害点落于高易发区和极高易发区,表明RF模型能更好地预测滑坡地质灾害的发生,耦合机器学习模型中,FR模型与RF模型的耦合效果同样也是优于SVM模型。

4种模型中极高易发区和高易发区占研究区总面积比例分别为25.58%、25.91%、22.44%、22.91%,4种模型的占比都较小,符合研究区的实

际情况,且耦合机器学习模型与单一机器学习模型相比,高易发区和极高易发区的面积占比均低于单一的机器学习模型,但是却能控制更多的滑坡地质灾害点,预测效果更佳。滑坡比率是不同易发性分区中滑坡地质灾害数目所占比例与易发分区所占比例的比值,是衡量易发分区是否合理准确的重要指标,统计显示,SVM、RF、FR-SVM、FR-RF模型的极高易发区的滑坡比例分别为4.4589、10.4369、7.6135以及9.3108,都有较高的滑坡比率,且随着易发性分区等级从极低易发区到极高易发区明显增大,符合滑坡易发性分区的结果。

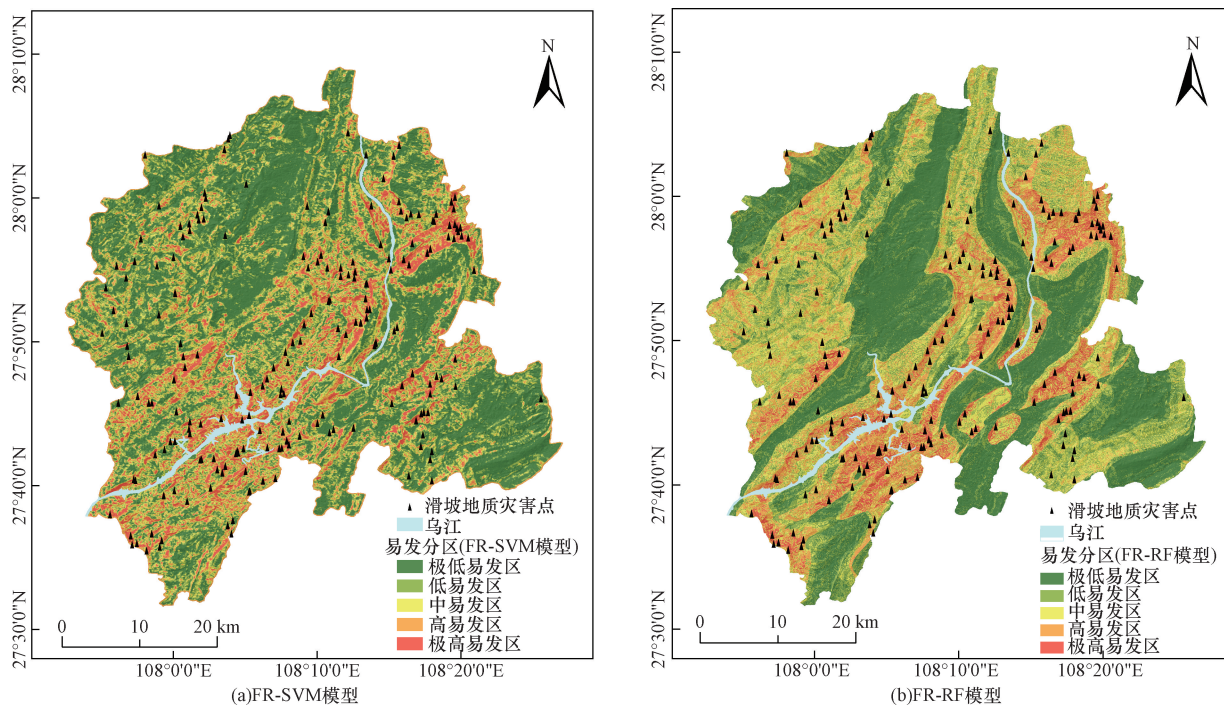


图 10 基于 FR-SVM 模型和 FR-RF 模型的滑坡易发性分区图

Fig. 10 Landslide susceptibility zoning map based on FR-SVM and FR-RF models

表 2 各模型易发性分区统计表

Table 2 Statistics table of landslide susceptibility zoning for each model

评价模型	易发性等级	滑坡数/次	分级栅格数量/个	占总滑坡数比例/%	占总栅格数比例/%	滑坡比例
SVM	极低易发区	4	853 916	2.00	34.76	0.057 5
	低易发区	15	564 384	7.50	22.98	0.326 4
	中易发区	31	409 862	15.50	16.69	0.928 9
	高易发区	48	347 300	24.00	14.14	1.697 5
	极高易发区	102	280 898	51.00	11.44	4.459 8
RF	极低易发区	0	268 356	0.00	10.92	0
	低易发区	4	740 121	2.00	30.13	0.066 4
	中易发区	22	811 411	11.00	33.03	0.333 0
	高易发区	75	519 972	37.50	21.17	1.771 5
	极高易发区	99	116 500	49.50	4.74	10.436 9
FR-SVM	极低易发区	4	916 438	2.00	37.31	0.053 6
	低易发区	8	586 568	4.00	23.88	0.167 5
	中易发区	18	402 272	9.00	16.38	0.549 6
	高易发区	52	360 729	26.00	14.69	1.770 5
	极高易发区	118	190 353	59.00	7.75	7.613 5
FR-RF	极低易发区	0	743 945	0.00	30.29	0
	低易发区	3	603 897	1.50	24.59	0.061 0
	中易发区	15	545 709	7.50	22.22	0.337 6
	高易发区	65	408 476	32.50	16.63	1.954 4
	极高易发区	117	154 333	58.50	6.28	9.310 8

使用 ROC 曲线进行不同滑坡易发性评价模型的精度验证,ROC 曲线的横坐标代表假阳性率(1-特异性),在滑坡易发性评价中代表被错误分类为正类别的负类别样本在总负类别样本中的比例,纵坐标代表真阳性率(敏感度),代表被正确预测为正类别的样本占总正类别样本的比例。为了更加准

地衡量评价结果,通常用 ROC 曲线下面积(AUC)衡量模型预测的准确度^[29]。滑坡易发性评价中 AUC 值范围通常在 0.5~1,其值越接近 1,模型的预测准确性越高,当 AUC<0.5 时,得到的结果是不符合客观事实的,AUC 范围在 0.5~0.7 说明模型精度较差,在 0.7~0.9 说明精度较高,在 0.9~1 说明精度

很好。将测试集样本和计算得到的滑坡易发性指数输入 SPSS 软件中进行 ROC 分析。如图 11 所示, SVM、RF、FR-SVM、FR-RF 模型的 AUC 分别为 0.831 6、0.843 9、0.864 4、0.910 4, 所有模型 AUC 均大于 0.8, 表明所有模型均有较好的预测效果, 其中 FR-RF 模型的 AUC 大于 0.9, 有很好的预测效果, 单一机器学习模型中, RF 模型的 AUC 高于 SVM 模型, 耦合机器学习模型中, FR-RF 模型的 AUC 高于 FR-SVM 模型, 且同一机器学习模型的耦合模型 AUC 高于单一的机器学习模型。

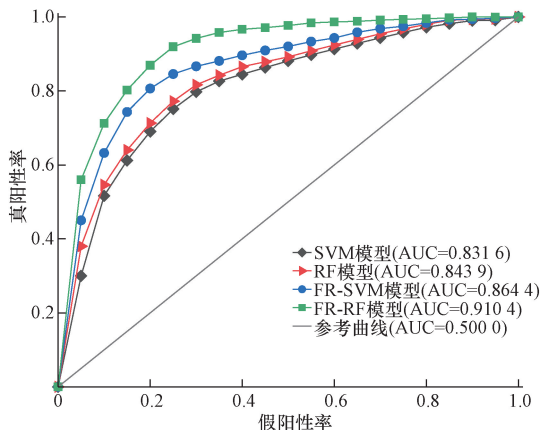


图 11 各模型 ROC 曲线

Fig. 11 ROC curves for each model

5 结论

(1) 以思南县为研究对象, 对研究区历史滑坡数据进行分析, 利用皮尔逊相关系数 (PCC) 和信息增益比 (IGR) 进行相关性分析确定滑坡易发性评价因子。通过 RF、SVM 模型及 FR 模型与 RF、SVM 模型耦合的 FR-RF、FR-SVM 模型开展研究区滑坡易发性评价。并且通过受试者工作曲线 (ROC) 下的曲线下面积 (AUC) 对上述模型进行精度评价, 结果表明 SVM、RF、FR-SVM、FR-RF 模型的 AUC 分别为 0.831 6、0.843 9、0.864 4、0.910 4, 4 个模型都有较好的预测精度。

(2) 为了解决不同评价因子间的数据不整合问题, 引入 FR 模型分别与 RF 模型和 SVM 进行耦合生成新的 FR-SVM 模型和 FR-RF 模型, 结果显示耦合机器学习模型与单一机器学习模型相比, 其 AUC 分别都高于单一机器学习模型, 从极高易发区和高易发区滑坡数占总滑坡数比例来看, SVM、RF、FR-SVM、FR-RF 模型极高易发区和高易发区滑坡数占比分别为 75.00%、87.00%、85.00%、91.00%, 耦合机器学习模型的滑坡数占比均是高于单一的机器学习模型, 耦合模型有着更好的预测精度。RF 模型和 FR-RF 模型的极低易发区均无滑坡灾害点, 表

明无论是单一机器学习模型还是耦合机器学习模型, RF 模型相比 SVM 模型都有着更好的预测精度。

(3) 研究区受喀斯特地貌影响, 极高易发区和高易发区主要分布在乌江水系沿岸以及地形起伏较大的区域 (香坝镇西北部、思林乡东北部、大河坝镇东部、关中坝街道南部、邵家桥镇东南部以及凉水井镇东南部), 同时也分布在道路以及城镇居民附近。与研究区所记录历史滑坡的实际情况相符合, 易发分区结果可为研究区滑坡灾害防范和减灾工作提供重要参考。

参考文献

- [1] Lin L, Lin Q, Wang Y. Landslide susceptibility mapping on a global scale using the method of logistic regression[J]. *Natural Hazards and Earth System Sciences*, 2017, 17(8): 1411-1424.
- [2] 杨豹, 赵瑞志, 王海波, 等. 遥感技术对地质灾害早期识别和动态监测——以昌波乡至羊拉乡段为例[J]. *科学技术与工程*, 2024, 24(5): 1823-1836.
Yang Bao, Zhao Ruizhi, Wang Haibo, et al. Remote sensing technology for early recognition and dynamic monitoring of geological disasters: take Changbo Township to Yangla Township section as an example[J]. *Science Technology and Engineering*, 2024, 24(5): 1823-1836.
- [3] Liu L, Zhang Y, Xiao T, et al. A frequency ratio-based sampling strategy for landslide susceptibility assessment[J]. *Bulletin of Engineering Geology and the Environment*, 2022, 81(9): 360.
- [4] Wang H, Zhang L, Luo H, et al. AI-powered landslide susceptibility assessment in Hong Kong[J]. *Engineering Geology*, 2021, 288(1): 106103.
- [5] 张泽方, 钱志宽, 魏勇, 等. 考虑最优影响因素组合的滑坡易发性评价——以水城区为例[J]. *科学技术与工程*, 2023, 23(10): 4091-4099.
Zhang Zefang, Qian Zhikuan, Wei Yong, et al. Landslide susceptibility evaluation considering optimal combination of influencing factors: a case study of Shuicheng District[J]. *Science Technology and Engineering*, 2023, 23(10): 4091-4099.
- [6] Deng H, Wu X, Zhang W, et al. Slope-unit scale landslide susceptibility mapping based on the random forest model in deep valley areas[J]. *Remote Sensing*, 2022, 14(17): 4252.
- [7] Zhao B, Ge Y, Chen H. Landslide susceptibility assessment for a transmission line in Gansu Province, China by using a hybrid approach of fractal theory, information value, and random forest models[J]. *Environmental Earth Sciences*, 2021, 80(12): 441.
- [8] Ozioko O H, Igwe O. GIS-based landslide susceptibility mapping using heuristic and bivariate statistical methods for Iva Valley and environs southeast Nigeria[J]. *Environmental Monitoring and Assessment*, 2020, 192(2): 119.
- [9] Ozturk D, Uzel-Gununi N. Investigation of the effects of hybrid modeling approaches, factor standardization, and categorical mapping on the performance of landslide susceptibility mapping in Van, Turkey[J]. *Natural Hazards*, 2022, 114(3): 2571-2604.
- [10] Panchal S, Shrivastava A K. A comparative study of frequency ratio, Shannon's entropy and analytic hierarchy process (AHP)

- models for landslide susceptibility assessment[J]. *ISPRS International Journal of Geo-Information*, 2021, 10(9): 603.
- [11] Gantimurova S, Parshin A, Erofeev V. GIS-based landslide susceptibility mapping of the Circum-Baikal Railway in Russia Using UAV data[J]. *Remote Sensing*, 2021, 13(18): 3629.
- [12] Xing X, Wu C, Li J, et al. Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method[J]. *Natural Hazards*, 2021, 106(1): 97-117.
- [13] Cheng J, Dai X, Wang Z, et al. Landslide susceptibility assessment model construction using typical machine learning for the Three Gorges Reservoir Area in China [J]. *Remote Sensing*, 2022, 14(9): 2257.
- [14] Yu L, Zhou C, Wang Y, et al. Coupling data-and knowledge-driven methods for landslide susceptibility mapping in human-modified environments: a case study from Wanzhou County, Three Gorges Reservoir Area, China[J]. *Remote Sensing*, 2022, 14(3): 774.
- [15] Chen W, Chen Y, Tsangartos P, et al. Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments[J]. *Remote Sensing*, 2020, 12(23): 3854.
- [16] Huang F, Yao C, Liu W, et al. Landslide susceptibility assessment in the Nantian area of China: a comparison of frequency ratio model and support vector machine[J]. *Geomatics, Natural Hazards and Risk*, 2018, 9(1): 919-938.
- [17] Fang Z, Wang Y, Duan H, 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.
- [18] Shou K, Lin J. Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale [J]. *Engineering Geology*, 2020, 265: 105434.
- [19] 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.
- [20] Luo X, Lin F, Zhu S, et al. Mine landslide susceptibility assessment using IVM, ANN and SVM models considering the contribution of affecting factors[J]. *Plos One*, 2019, 14(4): e0215134.
- [21] 刘亚静, 刘红健. 基于信息量-随机森林模型的地震带地质灾害易发性评价: 以松潘-较场地震带为例[J]. *科学技术与工程*, 2024, 24(1): 143-154.
- Liu Yajing, Liu Hongjian. Evaluation of geological hazard susceptibility in seismic zone based on information data-RF model: a case study of Songpan-Jiaochang seismic zone[J]. *Science Technology and Engineering*, 2024, 24(1): 143-154.
- [22] 何万才, 赵俊三, 林伊琳, 等. 基于证据权和支持向量机模型的威信县滑坡易发性评价[J]. *科学技术与工程*, 2023, 23(15): 6350-6360.
- He Wancai, Zhao Junsan, Lin Yilin, et al. Landslide susceptibility assessment in Weixin County based on evidence weight and support vector machine model[J]. *Science Technology and Engineering*, 2023, 23(15): 6350-6360.
- [23] 罗红, 陈磊, 姜运力, 等. 自然保护地整合优化的景观格局变化分析——以贵州省思南县为例[J]. *生态学报*, 2021, 41(20): 8076-8086.
- Luo Hong, Chen Lei, Jiang Yunli, et al. Landscape pattern changes and analysis for the integration and optimization of natural protected areas: a case study on Sinan County of Guizhou Province [J]. *Acta Ecologica Sinica*, 2021, 41(20): 8076-8086.
- [24] 李军, 周成虎. 基于栅格 GIS 滑坡风险评价方法中格网大小选取分析[J]. *遥感学报*, 2003(2): 86-92.
- Li Jun, Zhou Chenghu. Appropriate grid size for terrain based landslide risk assessment method based on grid GIS[J]. *Journal of Remote Sensing*, 2003(2): 86-92.
- [25] 张纪恺, 凌斯祥, 李晓宁, 等. 九寨沟县滑坡灾害易发性快速评估模型对比研究[J]. *岩石力学与工程学报*, 2020, 39(8): 1595-1610.
- Zhang Qikai, Ling Sixiang, Li Xiaoning, et al. Comparative study of rapid assessment models for landslide hazard susceptibility in Jiuzhaigou County [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2020, 39(8): 1595-1610.
- [26] 方迎潮, 王小松, 蒋毅, 等. 基于时序 InSAR 技术的中贵天然气管道天水市段沿线滑坡隐患识别与形变分析[J]. *科学技术与工程*, 2024, 24(4): 1357-1369.
- Fang Yingchao, Wang Xiaosong, Jiang Yi, et al. Landslide hazard identification and deformation analysis along the Tianshui City section of the China-Guizhou natural gas pipeline based on time-series InSAR technology[J]. *Science Technology and Engineering*, 2024, 24(4): 1357-1369.
- [27] Xie W, Nie W, Saffari P, et al. Landslide hazard assessment based on Bayesian optimization-support vector machine in Nanping City, China[J]. *Natural Hazards*, 2021, 109(1): 931-948.
- [28] Huang W, Ding M, Li Z, 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.
- [29] Sun X, Chen J, Han X, et al. Application of a GIS-based slope unit method for landslide susceptibility mapping along the rapidly uplifting section of the upper Jinsha River, south-western China [J]. *Bulletin of Engineering Geology and the Environment*, 2020, 79(1): 533-549.