

基于CF与优化RF模型耦合的泰山地区 地质灾害易发性评价

咸利民¹, 季民¹, 刘法军², 李强³

(1.山东科技大学 测绘与空间信息学院, 山东 青岛 266590; 2.山东省地质矿产勘查
开发局第五地质大队, 山东 泰安 271000; 3.聊城市地质矿产调查监测中心, 山东 聊城 252000)

摘要: [目的] 针对泰山地区地质灾害频发这一现状, 研究并构建地质灾害易发性评价模型, 为该地区的地质灾害预防与治理工作提供参考。[方法] 以泰山地区为研究区, 采用确定性系数模型与粒子群算法优化RF模型耦合的方法, 完成对研究区的地质灾害易发性评价。该方法是利用确定性系数(CF)模型计算影响因子对地质灾害的敏感值, 作为模型训练的属性值, 引入粒子群算法对随机森林(RF)模型进行参数寻优, 提高模型对地质灾害的预测精度和准确度。选取坡度、距道路距离、土地利用类型、植被指数等11个影响因子, 采用皮尔逊相关系数法和多重共线性检查进行影响因子筛选择优, 绘制ROC和PR曲线对训练模型进行精度评价。[结果] CF-PSO-RF耦合模型相比单一SVR、单一RF和CF-PSO-SVR模型的极高易发区面积比例分别提高10.55%、10.04%和5.08%, AUC值分别提高14%、5.1%和1.7%, AP精度分别提高了11.7%、4.4%、1.2%。预测结果显示, 泰山地区的极高、高易发区主要位于泰山景区、岱岳区北部等地形起伏和坡度较大的区域, 面积所占比例为28.05%, 涵盖了60.1%的地质灾害点; 相反, 低、极低易发区主要分布在建设用地、农田等地势平坦区域, 面积比例为59.26%。[结论] 将确定性系数模型与优化后RF模型耦合, 相比单一模型精度有进一步的提升, 又优于CF-PSO-SVR模型精度, 评价结果符合实际情况。

关键词: 地质灾害易发性评价; 粒子群算法; 确定性系数模型(CF); 随机森林模型(RF); 支持向量机模型(SVM); 泰山地区

文献标识码: B

文章编号: 1000-288X(2024)05-0134-10

中图分类号: P642.2, X43

文献参数: 咸利民, 季民, 刘法军, 等. 基于CF与优化RF模型耦合的泰山地区地质灾害易发性评价[J]. 水土保持通报, 2024, 44(5): 134-143. DOI: 10.13961/j.cnki.stbctb.2024.05.015; Xian Limin, Ji Min, Liu Fajun, et al. Assessing geological disaster susceptibility in Taishan area by coupling certainty factor model with optimized random forest model [J]. Bulletin of Soil and Water Conservation, 2024, 44(5): 134-143.

Assessing Geological Disaster Susceptibility in Taishan Area by Coupling Certainty Factor Model with Optimized Random Forest Model

Xian Limin¹, Ji Min¹, Liu Fajun², Li Qiang³

(1.College of Geodesy and Geomatics, Shandong University of Science and
Technology, Qingdao, Shandong 266590, China; 2.No.5 Institute of Geology and Mineral
Resources Exploration of Shandong Province, Tai'an, Shandong 271000, China; 3.The Geological and
Mineral Resources Survey and Monitoring Center of Liaocheng City, Liaocheng, Shandong 252000, China)

Abstract: [Objective] The present study was performed to develop a geological disaster susceptibility evaluation model for predicting frequent geological disasters in the Taishan area. The aim was to use the results of this susceptibility evaluation as reference for preventing and managing geological disasters in this area. [Methods] Our analyses focused on the Taishan area, employing a method in which the certainty factor (CF) model was coupled with a random forest (RF) model optimized using the particle swarm optimization (PSO) algorithm to evaluate the geological disaster susceptibility in the research area. This method uses the CF model to calculate the sensitivity values of the factors influencing geological disasters, which are then

收稿日期: 2024-06-07

修回日期: 2024-07-11

资助项目: 2023年山东省本科教学改革研究重大项目“面向智慧城市的测绘类专业人才培养探索与实践”(Z2023208)

第一作者: 咸利民(1999—), 男(汉族), 河北省张家口市人, 硕士研究生, 研究方向为地质灾害识别与评价。Email: 2028655392@qq.com。

通信作者: 季民(1970—), 男(汉族), 山东省德州市人, 博士, 教授, 主要从事自然资源监测与评价等方面的研究。Email: Jamesjimin@126.com。

used as attribute values for model training. The PSO algorithm was introduced to optimize the parameters of the RF model, thereby improving the accuracy and precision of the model in predicting geological disasters. Eleven influencing factors, including slope, distance to roads, land-use type, and vegetation index, were selected. The Pearson correlation coefficient method and a multicollinearity check were used to screen and optimize these influencing factors. The precision of the trained model was evaluated using ROC and PR curves. [Results] Compared with those of the single SVR, single RF, and CF-PSO-SVR models, the CF-PSO-RF coupled model significantly improved the proportion of extremely-high-susceptibility areas by 10.55%, 10.04%, and 5.08%, respectively, increased the AUC values by 14%, 5.1%, and 1.7%, respectively, and enhanced the average precision (AP) accuracy by 11.7%, 4.4%, and 1.2%, respectively. The prediction results revealed that the regions with high and extremely high susceptibility to geological disasters were mainly located in the Taishan scenic area, Northern Daiyue District, and other regions with significant topographic relief and steep slopes, covering 28.05% of the area and encompassing 60.1% of the geological disaster points. In contrast, regions with low and very low susceptibility were primarily distributed in flat areas, such as construction and farm lands, accounting for 59.26% of the total area. [Conclusion] The accuracy of the CF-PSO-RF coupled model for disaster susceptibility evaluation was notably higher than that of the single models; further, its precision was superior to that of the CF-PSO-SVR model. These evaluation results are consistent with actual conditions.

Keywords: geological hazard susceptibility assessment; particle swarm optimization algorithm; certainty factor (CF) model; random forest (RF) model; support vector machine (SVM) model; Taishan area

泰山地区地质构造复杂、地形地貌多样,是地质灾害的多发区^[1],截至2020年底,泰安市内地质灾害造成经济损失达 2.53×10^7 元,严重影响了当地居民的生活和经济发展。根据《山东省泰安市地质灾害防治规划(2021—2025年)》^[2],开展泰山地区的地质灾害易发性评价,对于防灾减灾工作意义重大。

目前地质灾害易发性评价模型主要分为定性评价模型和定量评价模型。定性评价模型包括层次分析法、模糊评价法和专家调查法等^[3],但其均具有较强的主观性且评价精度不高的缺点^[4-5];随着大数据技术和3S技术的发展,定量评价模型经历了数量统计模型和机器学习模型两个阶段^[6]。确定性系数模型(certainty factor, CF)作为典型的数理统计模型,能够反映出评价因子在不同分级类别下对地质灾害的敏感程度^[7],但忽视了各因子对易发性影响存在的差异性,难以准确反映各因子与灾害点之间的关系^[8]。机器学习模型能够在较小样本下处理复杂的非线性关系^[9],如支持向量机(support vector machines, SVM)^[10]、随机森林(random forest, RF)^[11]、逻辑回归(logistic regression, LR)^[12]、人工神经网络(artificial neural network, ANN)等^[13-14]模型,但也存在样本数据量纲不统一的问题,为此诸多学者采用多模型耦合的方法来提升评价精度。如郑迎凯等^[15]利用确定性系数模型与随机森林模型耦合的方法,对云南省芒市进行滑坡预测易发性评价,模

型精度达到91%,优于随机抽样的结果。为提升机器学习模型精度,超参数设置的合理性显得尤为重要^[16]。粒子群优化算法(PSO)具有全局搜索能力和快速收敛的特点,适用于复杂、高维的优化问题,在地质灾害的易发性评价中有良好的效果^[17]。如郭天颂^[18]基于粒子群优化算法,构建了PSO-SVR模型,对三峡区巫山段区域进行滑坡易发性评价,结果与历史滑坡有较好的一致性。本文在对泰山地区11个地质灾害影响因子进行择优选取的基础上,采用CF模型与PSO优化模型耦合的方法,分别构建CF-PSO-RF模型和CF-PSO-SVR模型,对泰山地区地质灾害的易发性进行评价,为该地区的地质灾害防治提供科学参考。

1 研究区概况

研究区为泰山地区,隶属山东省泰安市,位于行政区泰山区和岱岳区,116.84°—117.48°E,36.18°—36.47°N区域之间。区内地形地貌复杂,地貌类型齐全,地形起伏较大,包括丘陵、平原、山地等复杂地形,整体地势北高南低,北部山脊和沟谷纵横交错。区内水系发育丰富,汶水从西向东贯穿境内,支流分布北密南疏。气候属大陆性半湿润季风气候,降雨季节性分布明显,春季(3—5月)降水较少,夏季(6—9月)汛期降水则较大,表现出春旱夏涝的特点。区内地质构造复杂,地层区划属华北地层区,鲁西地层分区,东部

山区的太古界泰山岩群出露较多。石炭、二叠纪地层和侏罗、白垩纪地层几乎无出露,仅少量第三纪地层露出。泰山和徂徕山一带的中低山区因受侵蚀、构造切割强烈和地形坡度较大等影响,是泥石流、崩塌和滑坡等地质灾害的高发区。泰山山前至大汶河沿岸为山前冲洪积平原区,隐伏碳酸盐岩岩溶较为发育,岩溶塌陷频繁发生。此外,不合理的人类工程活动如采矿、开山、道路建设及土地开发,改变了原有的水文及地质条件,破坏了地表的自然形态和结构,导致水土流失严重,增加了泥石流、滑坡、崩塌等地质灾害发生的风险,对当地居民生命财产安全造成严重威胁,阻碍了经济发展。

2 研究方法 with 数据来源

2.1 数据来源

地质灾害的发生受到自然环境和人为活动的综合影响,本文主要选取的地质灾害影响因子分为地形地貌、水文环境、地质环境、人类工程 4 个方面^[19],其中部分地质灾害点由山东省地质矿产勘查开发局第五地质大队提供,其余地质灾害隐患点由野外验证和多源遥感技术(InSAR, RS)获得,其他数据来源详见表 1。为方便统计与分析,将整个研究区按照 30 m × 30 m 的栅格单元进行划分,共划分为 2 326 167 个评价单元。

表 1 数据来源

Table 1 Research data sources

分类	影响因子	数据来源	数据类型
地形地貌	坡度	美国国家航空航天局 NASA STRM 30 m DEM Landsat 30 m, 1986—2022 年最大 NDVI 数据集	30 × 30 栅格数据
	坡向		
	坡面曲率		
	地形起伏度		
	地形粗糙程度		
	归一化植被指数		
水文环境	距水流距离	国家基础地理信息中心 2021 年水系数据	矢量数据
地质环境	距断层距离	地震活动断层探测数据中心 2022 年中国区域主要断层数据	矢量数据
人类工程	距道路距离	国家基础地理信息中心 2021 年道路数据	矢量数据
	土地利用类型	GlobeLand 30, 2020 年土地利用数据集	30 × 30 栅格数据

2.2 确定性系数模型(CF)

确定性系数是一种概率函数^[20],可以计算各个影响因子在不同分级情况下对地质灾害的敏感性,CF 模型的表达式为:

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

式中: PP_a 为影响因子 a 在分级中的地质灾害发生的条件概率,表示为在影响因子 a 单元中存在的地质灾害隐患点的个数与该级单元面积的比值; PP_s 为整个研究区的地质灾害隐患的先验概率,表示为整个研究区内地质灾害隐患点数量与研究区总面积的比值。CF 取值介于 -1~1 之间, $CF > 0$ 时,值越接近 1,表明地质灾害发生有较大的可能性; $CF < 0$ 时,值越接近 -1,表明地质灾害发生有较小的可能性; $CF = 0$ 时,则无法判断该影响因子在此分级下地质灾害发生的可能性。

2.3 随机森林模型(RF)

随机森林模型是一种集成学习方法^[21],能够有

效处理高维数据和复杂的非线性问题。其回归思路为:从训练数据集中随机有放回的选取 K 次数据作为子数据集,随机选取部分特征作为输入,建立 K 个决策树模型,每个决策树单独工作输出一个结果,通过对所有决策树预测值的平均值作为最终结果^[22]。

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K T_k(x) \quad (2)$$

式中: K 为树的数量; $T_k(x)$ 为第 k 棵树对 x 的预测值。

2.4 支持向量机回归模型(SVR)

支持向量机回归(SVR)是支持向量机的一个回归分支,适用于高维数据特征、小样本数据等情况^[23]。其进行地质灾害预测的思路是:将每个样本的所有影响因子与对应的地质灾害发生概率组合为训练数据,通过非线性映射在低维空间和高维空间之间建立映射关系,构建回归预测函数 $f(x)$ 。

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad (3)$$

式中: a_i, a_i^* 为 Lagrange 系数; $K(x_i, x)$ 为核函数,本次核函数为径向基核函数(radial basis function)。

2.5 CF 与 PSO 优化模型耦合

粒子群算法具有收敛速度快、参数少、算法简单易实现的优点^[24],具备很好的全局搜索能力,本文引入粒子群算法对 RF 和 SVR 模型的超参数自适应寻优,最后将优化后的模型应用于易发性评价当中^[25]。

2.5.1 PSO 优化 RF 和 SVR 模型 通过初始化粒子群和 RF,SVR 模型的参数,计算适应度不断更新每个粒子的个体最优位置(p_best)和全局最优位置(g_best),寻找 RF 模型的树的数量($n_estimators$),树的最大深度(max_depth),每个节点的最小样本数($min_samples_split$)和 SVR 模型的惩罚参数(C)和径向基函数核(RBD Kernel)的最优参数,最后将粒子群优化算法优化后的最优超参数组合分别对应赋予 RF 模型和 SVR 模型并进行模型训练。其中,

速度更新公式为:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 [p_i^{best} - x_i(t)] + c_2 r_2 [g^{best} - x_i(t)] \quad (4)$$

式中: ω 是惯性权重; c_1, c_2 为学习因子; r_1, r_2 为随机数。

位置更新公式为:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

式中: x_i 为粒子位置; v_i 为粒子速度。

2.5.2 CF 与 PSO 优化模型耦合方法 鉴于 CF 模型能够消除各评价因子量纲差异造成的误差以及 RF 模型、SVR 模型能够客观反映评价因子与地质灾害之间关系的优点,利用 PSO 算法对模型进行参数寻优,提升模型的适应性和预测精度,CF 模型与 PSO 优化模型耦合过程如下:根据研究区地质灾害位置,建立 500 m 缓冲区随机选取等量的非地质灾害点,组成地质灾害样本集,将基于确定性系数模型计算得到的影响因子 CF 值作为模型训练样本数据和研究区的属性值,将训练样本数据按照 7:3 的比例划分训练集和验证集,利用 PSO 算法分别对 RF 模型、SVR 模型进行参数寻优,构建 CF-PSO-RF,CF-PSO-SVR 耦合模型,提取整个研究区的属性值并输入到耦合模型中,通过模型训练计算研究区地质灾害易发性指

数,预测地质灾害发生概率,完成对研究区的易发性评价。

3 影响因子评价指标体系建立

3.1 影响因子选取

根据研究区的地质构造、灾害发育特征、环境因素和人类活动,初步选取了坡度、坡向、坡面曲率、地形起伏度、地形粗糙程度、高程、距水流距离、距断层距离、归一化植被指数、距道路距离、土地利用类型 11 个影响因子进行易发性评价(图 1)^[26]。

3.2 影响因子筛选择优

皮尔逊相关系数法是对两个及以上的处于同等地位的随机变量之间的相关程度进行研究的统计分析方法,皮尔逊相关系数绝对值小于 0.5,表明各评价指标间没有显著的相关性;绝对值大于 0.5,表明各评价指标之间存在一定的相关性^[27]。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (6)$$

式中: r 为皮尔逊相关系数; \bar{X}, \bar{Y} 为 X 和 Y 的均值。

由图 2 可以看出,坡度与地形起伏度、地形粗糙度的相关系数分别为 0.96 和 0.97,远大于 0.5,说明 3 者之间具有极高的相关性,而且影响因子之间可能会存在共线性,造成模型复杂化、过拟合等问题,影响模型的评估精度^[28]。

利用方差膨胀因子 VIF 和容忍度 TOL 对影响因子进行共线性检查,当 $VIF > 2$ 或 $TOL < 0.5$ 时,表明因子间存在较强的多重共线性;反之,则因子间多重共线性较轻或不存在多重共线性^[29]。由表 2 可知,地形起伏度、地形粗糙度、坡度三者的 VIF 均大于 2,且容忍度 < 0.5 ,说明 3 者存在较强的共线性,结合其相关性结果与分析,坡度在不同程度上包含了地形起伏度和地形粗糙度的信息,可以反映二者的变化情况,因此,剔除地形起伏度和地形粗糙程度,保留其余 9 个影响因子建立评价指标体系。

表 2 影响因子共线性分析

Table 1 Collinearity analysis of influencing factors

因子编号	因子名称	容忍度	VIF	因子编号	因子名称	容忍度	VIF
1	坡面曲率	0.815	1.226	7	距水流距离	0.923	1.083
2	地形起伏度	0.074	13.532	8	高程分布	0.725	1.380
3	地形粗糙度	0.053	18.831	9	坡度	0.037	26.688
4	NDVI	0.778	1.286	10	坡向	0.901	1.110
5	距道路距离	0.895	1.117	11	土地利用类型	0.800	1.251
6	距断层距离	0.857	1.168				

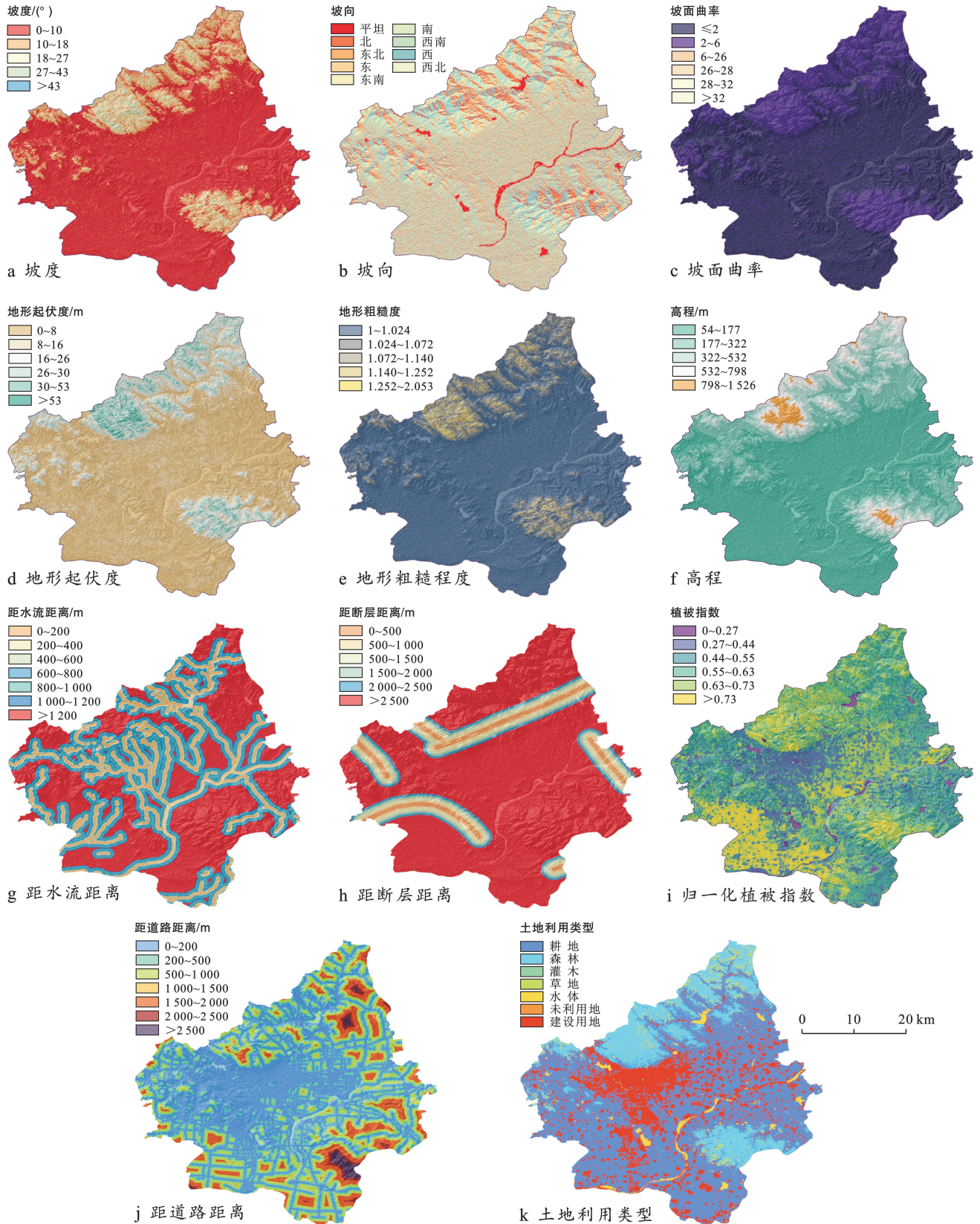


图 1 泰山区域地质灾害影响因子

Fig.1 Atlas of Taishan area geological hazards impact factors

3.3 评价指标确定性系数计算

对地质灾害影响因子进行分级处理,统计地质灾

害点在各个因子分级类别下的分布情况,利用确定性系数模型计算确定性系数值(表 3)。

4 结果与分析

4.1 易发性评价结果与分析

根据多个模型对地质灾害发生预测结果,采用自然间断分级法对研究区的地质灾害易发性程度划分为 5 个区间:极低易发区、低易发区、中等易发区、高易发区、极高易发区,同时利用均方误差、决定系数 R^2 , ROC 和 PR 曲线对模型训练结果进行精度验证。

(1) 从易发性评价结果区划图 3 可以看出,在研究区北部的高、极高易发性区域划分上,耦合模型相比于单一模型的预测结果分布更为聚集,且有更好的区分和过渡效果,而单一模型的预测结果分布较为分散,没有形成较为集中的风险区域;在研究区南部的低、极低易发性区域划分上,单一模型的预测结果相比于耦合模型较为模糊,预测结果呈“条带分布”,与实际情况不符;耦合模型中,CF-PSO-SVR 模型相较

于 CF-PSO-RF 模型在北部的极高易发性区域预测效果上较为保守,可能遗漏部分极高风险区域。

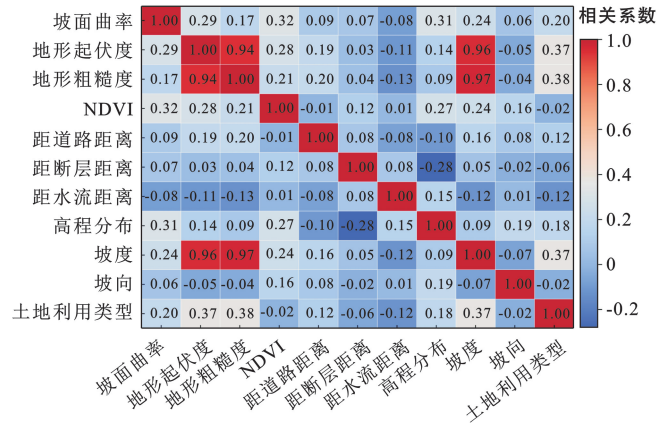


图 2 泰山地区地质灾害影响因子相关性系数矩阵热力图
Fig.2 Heat map of correlation coefficient matrix of geological disaster influencing factors in Taishan area

表 3 各影响因子分级情况及确定性系数(CF)

Table 3 Classification of various influencing factors and certainty coefficient (CF)

影响因子	编号	分级	CF	影响因子	编号	分级	CF
坡 向	1	平坦	-0.709	土地利用类型	1	耕地	-0.329
	2	北	-0.163		2	森林	0.624
	3	东北	-0.183		3	灌木	0.997
	4	东	-0.259		4	草地	0.858
	5	东南	0.512		5	水体	-1
	6	南	0.287		6	未利用地	-1
	7	西南	-0.157		7	建设用地	-0.959
	高 程	8	西	-0.416	1	[0,200)	0.153
		9	西北	-0.401	2	[200,400)	-0.124
1		[54,177)	-1	3	[400,600)	-0.244	
2		[177,322)	0.221	4	[600,800)	0.301	
3		[322,532)	0.787	5	[800,1 000)	0.201	
距道路距离	4	[532,798)	0.710	6	[1 000,1 200)	0.338	
	5	[798,1526)	0.520	7	≥1 200	0.040	
	1	[0,200)	0.128	距断层距离	1	[0,500)	-0.579
	2	[200,500)	0.016		2	[500,1 000)	-0.190
	3	[500,1 000)	-0.157		3	[1 000,1 500)	0.161
	4	[1 000,1 500)	-0.392		4	[1 500,2 000)	-0.615
	5	[1 500,2 000)	-0.354		5	[2 000,2 500)	-0.066
6	[2 000,2 500)	-0.171	6		≥2 500	0.069	
坡 度	7	≥2 500	0.066	坡度曲率	1	[1,2)	-0.525
	1	[0,10)	-0.730		2	[2,6)	0.633
	2	[10,18)	0.702		3	[6,26)	-1
	3	[18,27)	0.780		4	[26,28)	-1
	4	[27,43)	0.482		5	[28,32)	-1
5	≥43	-1	6		[32,39)	-1	
NDVI	1	[0,0.27)	-1	NDVI	4	[0.55,0.63)	0.310
	2	[0.27,0.44)	-0.844		5	[0.63,0.73)	0.118
	3	[0.44,0.55)	-0.104		6	≥0.73	-0.217

(2) 由图 3 可知,极高易发性区域主要集中分布在泰山区景区和岱岳区北部、西南部地区,地形起伏较大和水系发达密集的区域,推断此处受地形变化和湿度影响,导致地质结构趋于不稳定状态,需重视滑

坡、崩塌等地质灾害的发生,保障当地居民的生命财产安全;极低易发区主要分布在研究区中部,该区域大部分为耕地、农田和人类建设用地,发生地质灾害的可能性小。

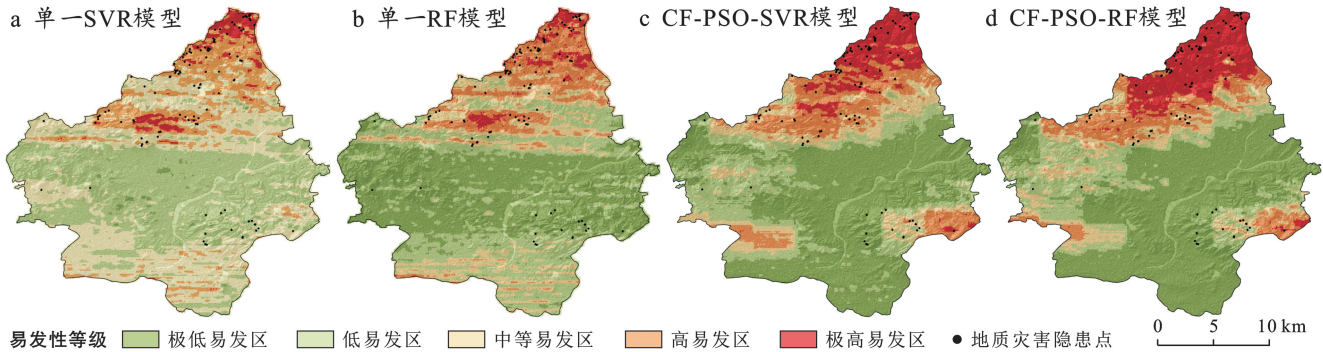


图 3 泰山地区地质灾害易发性评价的区划结果

Fig.3 Zoning results of geological disaster susceptibility evaluation in Taishan area

(3) 从易发性分区统计表 4 可以看出,单一模型的地质灾害在极低、低易发性区域面积比例相对耦合模型较高,在极高、高易发性面积比例较低,同时,在高、极高易发区中灾害点的数量比例相对较少,相反,在低、极低易发区的灾害点数量比例相对较高,

说明预测效果较差;但耦合模型中,CF-PSO-RF 模型易发性评价结果中极高易发区面积比例相对 CF-PSO-SVR 模型更大,在极高易发区内包含了 49.1% 的更多地质灾害点,说明 CF-PSO-RF 的预测结果更好。

表 4 4 种模型的泰山地区地质灾害易发性分区面积及隐患比例测算

Table 4 Zoning area and hidden danger proportion of geological disaster susceptibility in Taishan area by using four model

训练模型	单一 SVR 模型		单一 RF 模型		CF-PSO-SVR		CF-PSO-RF	
	面积比例/%	灾害点比例/%	面积比例/%	灾害点比例/%	面积比例/%	灾害点比例/%	面积比例/%	灾害点比例/%
极低易发区	2.78	0	43.08	11.8	46.28	5.5	42.48	7.2
低易发区	43.28	23.6	27.07	15.5	18.15	9.1	16.78	6.4
中等易发区	32.88	34.5	11.57	12.7	10.15	12.7	12.70	17.3
高易发区	16.41	25.5	13.13	35.5	15.30	30.0	12.85	20.0
极高易发区	4.65	16.4	5.16	24.5	10.12	42.7	15.20	49.1

4.2 模型精度对比

为了评估耦合模型与单一模型在地质灾害易发性预测中的精度,本文采用了 ROC(受试者工作特征曲线)^[30] 和 PR (precision-recall curve)^[31]。通过 ROC 曲线下的面积(AUC)来评估模型,横轴为假阳性率(特异性),表示误将非地质灾害点预测为灾害点的比例,纵轴为真阳性率(敏感性),表示正确预测为地质灾害点的比例。

PR 曲线则通过计算其面积下的平均精度(AP),横轴为召回率,表示所有实际灾害点中被正确预测的比例,纵轴为精确率,表示所有被预测为灾害点的样本中实际为灾害点的比例。

由图 4 可知,单独 SVR 模型和单独 RF 模型的

ROC 曲线 AUC 分别为 0.832 和 0.921,PR 曲线显示预测精度分别为 85.3% 和 92.6%,CF-PSO-SVR 模型和 CF-PSO-RF 模型的 AUC 分别为 0.955 和 0.972,PR 曲线显示预测精度分别为 95.8% 和 97.0%,相比对应的单一 SVR 模型和 RF 模型的 AUC 值提高 12.3% 和 5.1%,预测精度相比提高 10.5% 和 4.4%,说明粒子群算法和确定性系数模型对训练模型预测精度和性能进行了提升,在耦合模型中,由图 4 和表 5 结果可知,相比 CF-PSO-SVR 模型,CF-PSO-RF 模型的 AUC 值、 R^2 值和预测精度更高且均方误差(MSE)更低,表明 CF-PSO-RF 模型对地质灾害的预测效果更好,精确度更高,更适用于研究的地质灾害易发性评价研究。

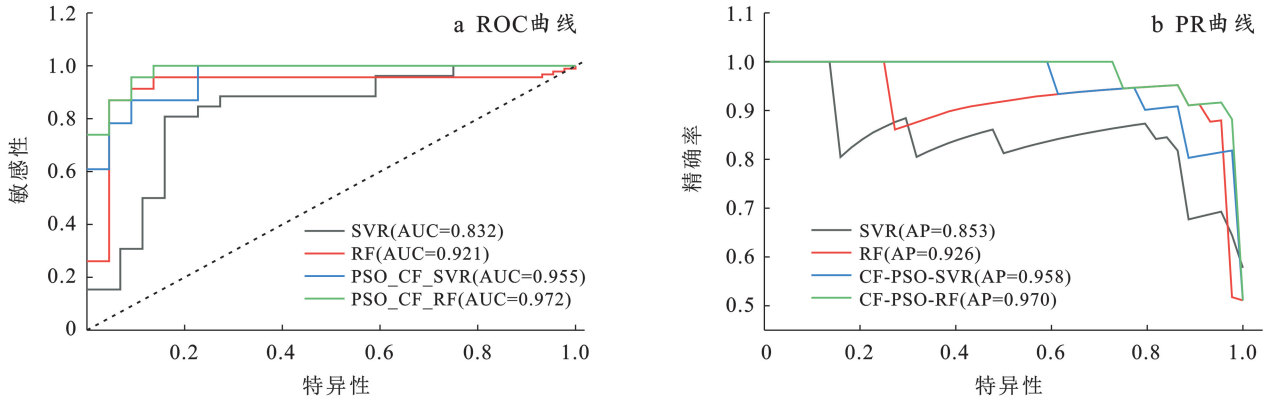


图 4 4 种模型的 ROC 和 PR 评价结果

Fig.4 Evaluation results of ROC and PR for four models

表 5 不同评价模型验证指标精度对比

Table 5 Comparison of accuracy of verification indicators for different evaluation models

训练模型	UC 值	平均精度 (AP)	R ²	均方误差 (MSE)
单一 SVR 模型	0.832	0.853	0.276 2	0.176 5
单一 RF 模型	0.921	0.926	0.576 2	0.202 4
CF-PSO-SVR	0.955	0.958	0.638 7	0.092 7
CF-PSO-RF	0.972	0.970	0.755 6	0.061 1

5 结论

(1) 从多角度分析 4 个模型可以得出,单一模型的易发性区域分布和各个精度指标均劣于耦合模型,其中 CF-PSO-RF 模型在各个精度指标和评价结果最优,说明利用 CF 模型对 PSO 算法优化后的机器学习模型进行耦合,能够消除评价因子因量纲差异导致的误差,并提升模型训练性能,进一步提高模型预测精度和评价结果的准确性。

(2) 基于 CF-PSO-RF 模型对泰山地区进行地质灾害易发性评价结果与实际地质灾害分布较为符合,其结果显示高、极高易发区主要分布在地、丘陵等起伏较大和碎岩裸露的区域,低、极低易发区则分布在平原、建设用地等地势平坦稳定的区域,评价结果与实际地质灾害情况吻合,可以为当地的地质灾害防治工作提供一定的参考。

本试验所建立的评价指标体系未考虑到土壤湿度、降水量等因素,以及各影响因子对地质灾害发生的贡献程度,可能会一定程度上对评价结果存在干扰,后续研究应充分考虑多种环境因素对地质灾害的影响,并利用主、客观赋值权重法结合获得影响因子权重,利用更新寻优算法(如麻雀寻优算法 SSA)提高评价模型预测精度。

参考文献 (References)

[1] 程凤,焦玉国,周超,等.泰安市崩塌、滑坡、泥石流发育特征及防治对策浅析[J].山东国土资源,2022,38(2):47-51.
Cheng Feng, Jiao Yuguo, Zhou Chao, et al. Analysis on development characteristics and prevention countermeasures of collapse landslide debris flow in Tai'an City [J]. Shandong Land and Resources, 2022,38(2):47-51.

[2] 泰安市自然资源和规划局.山东省泰安市地质灾害防治规划(2021-2025)[EB/OL].(2022-10-26). https://zrzyj.taian.gov.cn/art/2022/10/26/art_165860_10307956.html.
Tai'an Natural Resources and Planning Bureau. Geological Disaster Prevention and Control Plan of Tai'an City, Shandong Province(2021-2025)[EB/OL].(2022-10-26). https://zrzyj.taian.gov.cn/art/2022/10/26/art_165860_10307956.html.

[3] 兰盈盈,郭昶成,朱云福.地质灾害易发性评价方法综述[J].地质与资源,2024,33(1):65-73.
Lan Yingying, Guo Changcheng, Zhu Yunfu. Review on the evaluation methods of geological hazard susceptibility [J]. Geology and Resources, 2024,33(1):65-73.

[4] 王哲,易发成.基于层次分析法的绵阳市地质灾害易发性评价[J].自然灾害学报,2009,18(1):14-23.
Wang Zhe, Yi Facheng. AHP-based evaluation of occurrence easiness of geological disasters in Mianyang City [J]. Journal of Natural Disasters, 2009,18(1):14-23.

[5] 牛全福.基于 GIS 的地质灾害风险评估方法研究:以“4·14”玉树地震为例[D].甘肃 兰州:兰州大学,2011.
Niu Quanfu. Study on the method of geological disaster risk assessmentbased on GISA case study in “4·14” Yushu earthquake[D]. Lanzhou, Gansu: Lanzhou University,2011.

[6] Ado M, Amitab K, Maji A K, et al. Landslide susceptibility mapping using machine learning: A literature survey [J]. Remote Sensing, 2022,14(13):3029.

- [7] 毛正君,张瑾鸽,仲佳鑫,等.基于确定性系数法的梯田型黄土滑坡隐患影响因素分析[J].水土保持通报,2023,43(2):183-192.
Mao Zhengjun, Zhang Jingge, Zhong Jiaxin, et al. Sensitivity analysis on factors influencing loess terrace landslide potential using certainty factor method [J]. Bulletin of Soil and Water Conservation, 2023,43(2):183-192.
- [8] 田春山,刘希林,汪佳.基于 CF 和 Logistic 回归模型的广东省地质灾害易发性评价[J].水文地质工程地质,2016,43(6):154-161.
Tian Chunshan, Liu Xilin, Wang Jia. Geohazard susceptibility assessment based on CF model and Logistic Regression models in Guangdong [J]. Hydrogeology & Engineering Geology, 2016,43(6):154-161.
- [9] 仇文岗,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.
- [10] Yao X, Tham L G, Dai F C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China [J]. Geomorphology, 2008,101(4):572-582.
- [11] 杜鹏,陈宁生,伍康林,等.基于随机森林模型的藏东南地区滑坡易发性评价及主控因素分析[J].成都理工大学学报(自然科学版),2024,51(2):328-344.
Du Peng, Chen Ningsheng, Wu Kanglin, et al. Evaluation of landslide susceptibility in southeast Tibet based on a random forest model [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 2024,51(2):328-344.
- [12] 胡涛,樊鑫,王硕,等.基于逻辑回归模型和 3S 技术的思南县滑坡易发性评价[J].地质科技通报,2020,39(2):113-121.
Hu Tao, Fan Xin, Wang Shuo, et al. Landslide susceptibility evaluation of Sinan County using logistics regression model and 3S technology [J]. Bulletin of Geological Science and Technology, 2020,39(2):113-121.
- [13] Li Jian. Study on Mine Geological Hazard Assessment Model Based on ANN [R]. 2010 2nd International Conference on Information Engineering and Computer Science. IEEE, 2010.
- [14] Tan Qulin, Huang Yong, Hu Jun, et al. Application of artificial neural network model based on GIS in geological hazard zoning [J]. Neural Computing and Applications, 2021,33(2):591-602.
- [15] 郑迎凯,陈建国,王成彬,等.确定性系数与随机森林模型在云南芒市滑坡易发性评价中的应用[J].地质科技通报,2020,39(6):131-144.
Zheng Yingkai, Chen Jianguo, Wang Chengbin, et al. Application of certainty factor and random forests model in landslide susceptibility evaluation in Mangshi City, Yunnan Province [J]. Bulletin of Geological Science and Technology, 2020,39(6):131-144.
- [16] 周晓亭,黄发明,吴伟成,等.基于耦合信息量法选择负样本的区域滑坡易发性预测[J].工程科学与技术,2022,54(3):25-35.
Zhou Xiaoting, Huang Faming, Wu Weicheng, et al. Regional landslide susceptibility prediction based on negative sample selected by coupling information value method [J]. Advanced Engineering Sciences, 2022,54(3):25-35.
- [17] 任瑞斌,李丽敏,王莲霞,等.基于 PSO-LSSVM 的广西花岗岩分布区滑坡易发性评价[J].国外电子测量技术,2023,42(5):157-162.
Ren Ruibin, Li Limin, Wang Lianxia, et al. Evaluation of landslide hazard susceptibility in granite distribution area of Guangxi based on PSO-LSSVM [J]. Foreign Electronic Measurement Technology, 2023,42(5):157-162.
- [18] 郭天颂.基于粒子群优化支持向量机的滑坡易发性评价与滑坡位移预测[D].陕西 西安:长安大学,2019.
Guo Tiansong. Landslide susceptibility evaluation and landslide displacement prediction based on particle swarm optimization: Based support vector machine [D]. Xi'an, Shaanxi: Chang'an University, 2019.
- [19] 陈芯宇,师芸,赵侃,等.基于 CF 融入 SSA 优化 SVM 和 RF 模型的滑坡易发性评价[J].西安理工大学学报,2024,40(1):121-131.
Chen Xinyu, Shi Yun, Zhao Kan, et al. Landslide susceptibility evaluation based on CF integrated with SSA to optimize SVM and RF models [J]. Journal of Xi'an University of Technology, 2024,40(1):121-131.
- [20] Shortliffe E H, Buchanan B G. A model of inexact reasoning in medicine [J]. Mathematical Biosciences, 1975,23(3/4):351-379.
- [21] Belgiu M, Drăgut L. Random forest in remote sensing: A review of applications and future directions [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016,114:24-31.
- [22] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al. A review of technologies on random forests [J]. Statistics & Information Forum, 2011,26(3):32-38.
- [23] 武雪玲,任福,牛瑞卿,等.斜坡单元支持下的滑坡易发性评价支持向量机模型[J].武汉大学学报(信息科学

- 版),2013,38(12):1499-1503.
- Wu Xueling, Ren Fu, Niu Ruiqing, et al. Landslide spatial prediction based on slope units and support vector machines [J]. *Geomatics and Information Science of Wuhan University*, 2013,38(12):1499-1503.
- [24] Clerc M, Kennedy J. The particle swarm - explosion, stability, and convergence in a multidimensional complex space [J]. *IEEE Transactions on Evolutionary Computation*, 2002,6(1):58-73.
- [25] 郭天颂,张菊清,韩煜,等.基于粒子群优化支持向量机的延长县滑坡易发性评价[J].*地质科技情报*,2019,38(3):236-243.
- Guo Tiansong, Zhang Juqing, Han Yu, et al. Evaluation of landslide susceptibility in Yanchang County based on particle swarm optimization-based support vector machine [J]. *Geological Science and Technology Information*, 2019,38(3):236-243.
- [26] 杨光,徐佩华,曹琛,等.基于确定性系数组合模型的区域滑坡敏感性评价[J].*工程地质学报*,2019,27(5):1153-1163.
- Yang Guang, Xu Peihua, Cao Chen, et al. Assessment of regional landslide susceptibility based on combined model of certainty factor method [J]. *Journal of Engineering Geology*, 2019,27(5):1153-1163.
- [27] 栗泽桐,王涛,周杨,等.基于信息量、逻辑回归及其耦合模型的滑坡易发性评估研究:以青海沙塘川流域为例[J].*现代地质*,2019,33(1):235-245.
- Li Zetong, Wang Tao, Zhou Yang, et al. Landslide susceptibility assessment based on information value model, logistic regression model and their integrated model: A case in Shatang River Basin, Qinghai Province [J]. *Geoscience*, 2019,33(1):235-245.
- [28] 黄发明,陈彬,毛达雄,等.基于自筛选深度学习的滑坡易发性预测建模及其可解释性[J].*地球科学*,2023,48(5):1696-1710.
- Huang Faming, Chen Bin, Mao Daxiong, et al. Landslide susceptibility prediction modeling and interpretability based on self-screening deep learning model [J]. *Earth Science*, 2023,48(5):1696-1710.
- [29] Chen Wei, Chen Yunzhi, Tsangaratos P, et al. Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments [J]. *Remote Sensing*, 2020,12(23):3854.
- [30] Lee S T, Yu T T, Peng W F, et al. Incorporating the effects of topographic amplification in the analysis of earthquake-induced landslide hazards using logistic regression [J]. *Natural Hazards and Earth System Sciences*, 2010,10(12):2475-2488.
- [31] 王丽霞,喜文飞,史正涛,等.深度学习滑坡识别算法中样本不平衡问题的研究[J].*测绘通报*,2024(5):12-18.
- Wang Lixia, Xi Wenfei, Shi Zhengtao, et al. Study on sample unbalance in landslide recognition algorithm based on depth learning [J]. *Bulletin of Surveying and Mapping*, 2024(5):12-18.
- ~~~~~
- (上接第 133 页)
- [59] 张妙仙,杨劲松.灌溉入渗条件下农田土壤水盐动态简化模型及应用[J].*土壤学报*,2002,39(1):75-82.
- Zhang Miaoxian, Yang Jinsong. The simplified model of salt-water regime in cropland soil under infiltration condition and its application [J]. *Acta Pedologica Sinica*, 2002,39(1):75-82.
- [60] 李玮,王立,姜涛.地下水浅埋区盐碱地滴灌条件下土壤盐分运移研究[J].*干旱地区农业研究*,2007,25(5):130-135.
- Li Wei, Wang Li, Jiang Tao. Progress of the study on salt transport in saline soil under drip irrigation condition in shallow groundwater area [J]. *Agricultural Research in the Arid Areas*, 2007,25(5):130-135.
- [61] 朱成立,徐雨琳,黄明逸,等.基于 AquaCrop 模型的冬小麦咸淡轮灌制度模拟与评价[J].*农业机械学报*,2022,53(4):330-342.
- Zhu Chengli, Xu Yulin, Huang Mingyi, et al. Simulation and evaluation of cycle irrigation with brackish and fresh water for winter wheat based on AquaCrop model [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022,53(4):330-342.
- [62] 黄亚捷,李贞,卓志清,等.用 SahysMod 模型研究不同灌排管理情景土壤水盐动态[J].*农业工程学报*,2020,36(11):129-140.
- Huang Yajie, Li Zhen, Zhuo Zhiqing, et al. Soil water and salt dynamics under different irrigation and drainage management scenarios based on SahysMod model [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020,36(11):129-140.