



基于信息量法和集成学习算法的地质灾害易发性评价 ——以黑龙江省哈尔滨市为例

李蕴峰, 卢彦达, 陈卓, 卢昱润, 李涛涛

中国地质调查局 哈尔滨自然资源综合调查中心, 黑龙江 哈尔滨 150000

摘要: 为开展黑龙江省哈尔滨市地质灾害易发性区划和地质灾害防治, 选取坡度、坡向、曲率、岩性、NDVI、距水系距离、距道路距离、距构造距离等 8 类评价因子, 建立地质灾害易发性评价指标体系。从信息量算法计算出的极低易发区和低易发区中随机选取非地质灾害样本, 与地质灾害样本组成论文数据集。采用随机森林、Adaboost 和 Stacking 等 3 种集成学习方法对哈尔滨市地质灾害易发性进行评价, 并通过混淆矩阵进行精度验证。结果表明: 4 种算法易发性评价分区图评价结果趋势相同, 且与研究区实际情况较为一致。哈尔滨市地质灾害主要诱发因素为人类工程活动, 极高发区主要集中在道路附近。随机森林算法预测的极高易发区的面积仅占全区的 1.27%, 地质灾害数量占比 21.03%, 频率比达 16.58, AUC 值为最高的 0.891, 说明 3 种集成学习算法中, 随机森林算法在该区域地质灾害易发性评价中更具优势。

关键词: 地质灾害; 易发性评价; 信息量法; 集成学习算法; 哈尔滨市

Assessment of geological hazard susceptibility based on information method and ensemble learning algorithm: A case study of Harbin City in Heilongjiang Province

LI Yun-feng, LU Yan-da, CHEN Zhuo, LU Yu-run, LI Tao-tao

Harbin Natural Resources Comprehensive Survey Center, CGS, Harbin 150000, China

Abstract: To carry out the geological hazard susceptibility zoning and prevention in Harbin, Heilongjiang Province, eight evaluation factors including slope gradient, slope aspect, curvature, lithology, NDVI, distance from river, distance from road and distance from structure are selected to establish the evaluation index system of geological hazard susceptibility. The non-geological hazard samples are randomly chosen from the extremely low and low susceptible zones calculated by information algorithm, which forms the document data set together with the geological hazard samples. Besides, three ensemble learning methods such as random forest (RF), Adaboost and Stacking are used to assess the geological hazard vulnerability in Harbin, and the accuracy is verified by confusion matrix. The results show that the trend of evaluation results of the four algorithms is the same, and consistent with the actual situation of the study area. The major inducing factor of geological hazards in Harbin is human engineering activities, with the extremely high

收稿日期: 2023-10-30; 修回日期: 2023-12-18. 编辑: 张哲.

基金项目: 中国地质调查局项目“应用地质调查数据应用服务”(DD20230595、DD20230594).

作者简介: 李蕴峰(1993—), 男, 工程师, 从事模型定量评估与评价, 通信地址 黑龙江省哈尔滨市南岗区保健副路 1 号, E-mail//liyunfeng9319@126.com

通信作者: 陈卓(1989—), 男, 博士, 高级工程师, 从事遥感矿产预测和环境评价, 通信地址 黑龙江省哈尔滨市南岗区保健副路 1 号, E-mail//chenz121@163.com

引用格式: 李蕴峰, 卢彦达, 陈卓, 卢昱润, 李涛涛. 基于信息量法和集成学习算法的地质灾害易发性评价——以黑龙江省哈尔滨市为例[J]. 地质与资源, 2025, 34(1): 77-86.

Li Y F, Lu Y D, Chen Z, Lu Y R, Li T T. Assessment of geological hazard susceptibility based on information method and ensemble learning algorithm: A case study of Harbin City in Heilongjiang Province [J]. Geology and Resources, 2025, 34(1): 77-86.

susceptible zones mainly concentrated near roads. The area of extremely high susceptible zones predicted by RF algorithm accounts for only 1.27% of the whole region, yet the number of geological hazards takes up 21.03%, with the frequency ratio of 16.58 and the maximum AUC value reaching 0.891, indicating that RF algorithm has more advantages in the geological hazard susceptibility evaluation among the above three algorithms.

Key words: geological hazard; susceptibility assessment; information method; ensemble learning algorithm; Harbin City

0 引言

地质灾害易发性区划是通过分析影响地质灾害的内因因素和外因因素,评价潜在地质灾害的空间地理位置分布^[1-2].近年来,大量学者采用多种评价方法进行地质灾害易发性评价研究.评价方法主要分为定性评价和定量评价,定性评价方法以知识驱动,主观性较强,主要包括专家打分法^[3]、层次分析法^[4]等;定量评价方法以数据驱动,利用概率统计建模,主要包括确定性系数法^[5]、信息量法^[6]、证据权法^[7]等.随着人工智能的发展,越来越多的学者采用统计分析算法与机器学习相结合的方法进行地质灾害易发性评价,取得了良好的应用效果^[8-11].机器学习拓展了地质灾害易发性评价方法选择,但是由于训练样本的有限性和地质灾害的复杂性,单个学习算法可能会错过假设空间中样本集的最佳拟合函数或真实分布,从而影响预测精度^[12].从多种机器学习方法获得的集成中值预测可提供更大的空间细节并产生最高的预测精度^[13],因此,集成学习预测方法比任何单一预测技术都更适合进行预测评价^[14].由于在性能增强上表现出的优越性,集成学习算法已开始应用于各个领域^[15],并取得了较好的效果.然而在地质灾害易发性评价中,集成学习却很少被使用.

随着哈尔滨市社会经济的快速发展,人类工程活动的增多,地质环境承受着巨大的影响和压力,严重影响着城市居民生命和财产安全.本文以黑龙江省哈尔滨市为研究区,选取随机森林、Adaboost和Stacking等3种集成方法对哈尔滨市地质灾害易发性进行评价.期望研究成果可为哈尔滨市有效开展地质灾害综合防治、保障经济社会可持续发展提供灾害风险信息 and 科学决策依据.

1 研究区概述

哈尔滨市位于黑龙江省西南部,地理坐标为东经125°42′—130°10′,北纬44°04′—46°40′,总

面积53 076.4×10⁸ km².海拔较低,地势较为平坦(图1).大地构造位于滨太平洋地层区之松辽分区的松嫩小区,松辽拗陷的中部东南端,东南隆起区边缘地带.研究区内新构造运动比较活跃,主要是继承和改造了燕山运动的构造格架和运动方式,以差异性升降为主要的运动方式^[16].区域内人类工程活动频繁,以道路和基础工程建设为主.哈尔滨市地处中纬度亚洲大陆东缘,冬季在内蒙古高压控制下,盛行西北风,寒冷干燥;夏季在太平洋副热带高压控制之下,盛行东南风,高温多雨.因此,研究区具有明显的季风气候特征,属中温带大陆性季风气候.哈尔滨市区内最大水系为松花江,是黑龙江右岸一大支流,为本区一级河流,境内长度达135 km,由南西向北东流经研究区.河道弯曲,河床宽度0.5~1.5 km,漫滩较宽,水深4~7 m.支流有拉林河、阿什河、呼兰河等,属区内二级河流.多年平均径流量376×10⁸ m³,年最大为846.7×10⁸ m³,年最小为122.5×10⁸ m³,年变差724.2×10⁸ m³.

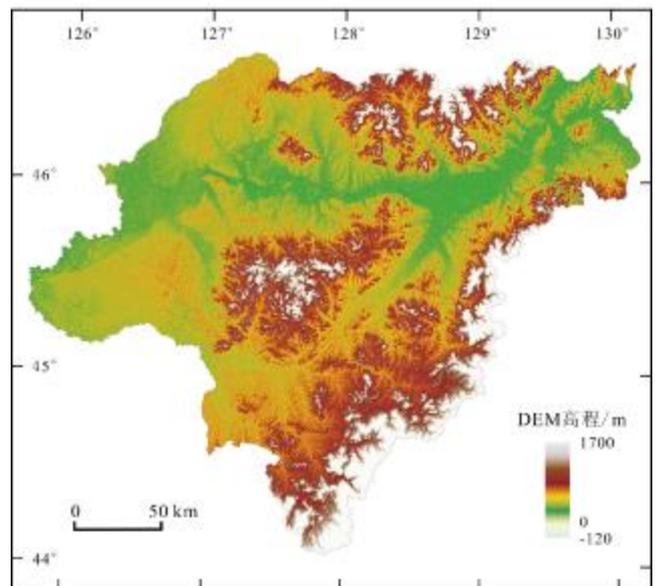


图1 研究区地形地貌图

Fig. 1 Topographic map of the study area

2 评价方法简介

2.1 信息量算法

信息量算法的理论基础是信息论,采用地质灾害发生过程中熵的减少来表征地质灾害事件发生的可能性^[17]。通常认为,信息量值大于零,表明该区域易发生地质灾害;信息量值小于零,表明该区域发生地质灾害的可能性低;信息量值等于零,表明该区域不发生地质灾害。计算公式如下:

$$I = \sum_{i=1}^n \log_2 \frac{S_{i0}/S_i}{A_0/A} \quad (1)$$

式中, I 为评价区某单元信息量预测值, S_i 为因素 x_i 所占单元总面积, S_{i0} 为因素 x_i 单元中发生地质灾害的单元面积之和, A 为区域内单元总面积, A_0 为已经发生地质灾害的单元面积之和。

2.2 Bagging 集成算法

Bagging 集成算法是并行式集成学习的最著名代表,其核心是通过分别构建多个并行的弱学习器,将多个弱学习器结果结合的团体学习算法。随机森林算法是基于树算法的 Bagging 的优化版本,是对样本和属性的抽样有放回的机器学习方法^[18]。随机森林算法是通过抽取不同的数据子集、不同的属性特征,构建多个决策树,而后对多个决策树的判断结果进行投票,得到随机森林的最终输出结果。

2.3 Boosting 集成算法

Boosting 集成算法是减少监督式学习中偏差的机器学习算法,其核心是通过弱学习器的不断集成,每轮迭代产生的新学习器,重点改进之前没处理好的地方,形成一个强学习器的集成学习算法。Adaboost 算法是 Boosting 算法系列的经典算法,是通过加权前轮基学习器分错样本,生成新的全体样本进行下轮基学习器训练,直到达到某预定的足够小的错误率或达到预先指定的最大迭代次数^[19]。

2.4 Stacking 集成算法

Stacking 集成算法是分成算法集成框架,其核心是通过基学习器综合估计所有基学习器的误差,利用元学习器减少预测残差^[20]。Stacking 集成算法可以理解将为基学习器结果作为数据特征,利用元学习器方法进行算法训练的过程。本研究使用两层 Stacking 集成学习方法,第一层选择随机森林、支持向量机和 Adaboost 算法作为基学习器。研究表明逻辑回归等结

构简单的算法更适合作为元学习器^[21],因此第二层选择逻辑回归算法作为元学习器。

3 地质灾害易发性评价指标体系

3.1 评价因子的选取

地质灾害的发生是内部地质条件因素和外部因素相互影响、共同作用的结果,合理选择影响因子是地质灾害易发性评价的关键^[22]。结合研究区自然地理特征、地质灾害分布特征,综合考量各类地质灾害相关数据的可获取性和精度,在参阅以往研究的基础上^[23-26],本研究初步选取高程、坡度、坡向、平面曲率、岩性、归一化植被指数(NDVI)、距水系距离、距道路距离、距构造距离等 9 类评价因子。

地质灾害数据来源于全球灾害数据平台(<https://www.gddat.cn/>),经统计,哈尔滨市崩塌和滑坡地质灾害共计 214 处。数字高程算法(DEM)数据来源于地理空间数据云平台(<https://www.gscloud.cn/>),下载数据分辨率为 30 m,坡度、坡向、曲率通过 DEM 数据和 ArcGIS 表面分析功能获取。研究区范围、道路、水系数据来源于国家地理信息公共服务平台(<https://www.tianditu.gov.cn/>),数据精度为 1:5 万。地质和构造数据来源于中国陆域 1:25 万分幅建造构造图空间数据库(V1)^[27]。NDVI 数据来源于国家生态科学数据中心(<http://www.nesdc.org.cn/>)^[28]。

考虑到因子间的相互耦合影响,为保证评价因子间的相互独立性和客观性,采用皮尔逊相关系数进行相关性程度分析。当 2 个因子的相关系数绝对值大于 0.4 时,认为存在明显相关性,应予以剔除。影响因子间的相关系数结果见表 1。

由表 1 可知,高程和坡度的皮尔逊相关系数为 0.605,同时与地层的相关系数为 0.565,因此剔除高程因子。最终选定坡度、坡向、曲率、岩性、NDVI、距水系距离、距道路距离、距构造距离等共计 8 类评价因子作为一级指标进行评价。

3.2 评价因子的分级及信息量计算

评价因子分级以各指标状态下一定步长的信息量与对应的分级地质灾害比例和分级栅格比例分布曲线为依据,对比两类曲线的分布形式,并将信息量分布曲线的突变点作为各因素状态等级划分的临界值^[17]。各因子分级范围及不同分级下的信息量计算结果如表 2。

表 1 影响因子间的相关系数
Table 1 Correlation coefficients of influencing factors

评价因子	高程	距水系距离	曲率	坡向	坡度	距道路距离	距构造距离	地层	NDVI
高程	1								
距水系距离	-0.119	1							
曲率	0.042	0	1						
坡向	0.045	-0.005	0	1					
坡度	0.605	-0.078	0.009	0.049	1				
距道路距离	-0.159	0.005	0	-0.004	-0.012	1			
距构造距离	0.092	-0.021	0.001	0.009	0.082	-0.024	1		
地层	0.565	-0.101	0.009	0.038	0.049	-0.17	0.111	1	
NDVI	0.357	-0.257	0	0.016	0.288	-0.136	0.073	0.362	1

表 2 地质灾害易发性评价指标信息量表
Table 2 Information value of geological hazard susceptibility assessment indexes

影响因素	二类评价因子	信息量值	排序	影响因素	二类评价因子	信息量值	排序
坡度/(°)	0-3	-0.525	33	NDVI	0-0.51	-0.811	38
	3-6	-0.185	29		0.51-0.73	0.726	11
	6-10	0.387	18		0.73-0.82	0.431	16
	10-16	0.724	12		0.82-0.87	0.443	15
	16-90	-0.116	27		>0.87	-0.801	37
坡向/(°)	0-73	-0.607	34	距水系距离/m	0-100	0.107	23
	73-146	-0.068	26		100-200	1.740	2
	146-212	0.765	9		200-300	0.764	10
	212-270	0.329	21		300-400	-0.242	30
	270-360	-0.611	35		>400	-0.026	24
曲率	<-0.78	1.049	6	距道路距离/m	0-100	2.734	1
	-0.78~-0.33	0.386	19		100-300	1.434	3
	-0.33-0	-0.172	28		300-500	0.367	20
	0-0.33	-0.495	31		500-700	0.675	14
	>0.33	0.869	7		>700	-0.834	39
岩性	第四系	-0.645	36	距构造距离/m	0-100	-1.049	40
	沉积岩	0.696	13		100-300	-0.497	32
	喷出岩	0.787	8		300-500	0.213	22
	侵入岩	0.410	17		500-800	1.346	4
	变质岩	1.295	5		>800	-0.037	25

根据因子分级范围,各因子分层结果可视化效果如图 2 所示.

4 地质灾害易发性评价与对比

4.1 训练与测试样本选取

结合收集资料的精度和研究区范围,本文选用 100 m 栅格进行数据计算,将研究区划分为 3 494 行和 2 984 列,共计 5 322 188 个栅格.研究区共有地质

灾害点 214 处,即 214 个正样本.负样本点的选取将直接影响算法拟合的效果.为有效避免因数据冗余造成的算法分类精度下降,采用信息量算法计算研究区各栅格信息量值,通过自然断点法将研究区划分为极低易发区、低易发区、中易发区、高易发区、极高易发区 5 类.根据信息量法计算出的各易发性等级比例,在极低易发区和低易发区采用 ArcGIS 中创建随机点工具,生成间隔 1 km 以上的非地质灾害点 1 750 个,作为负

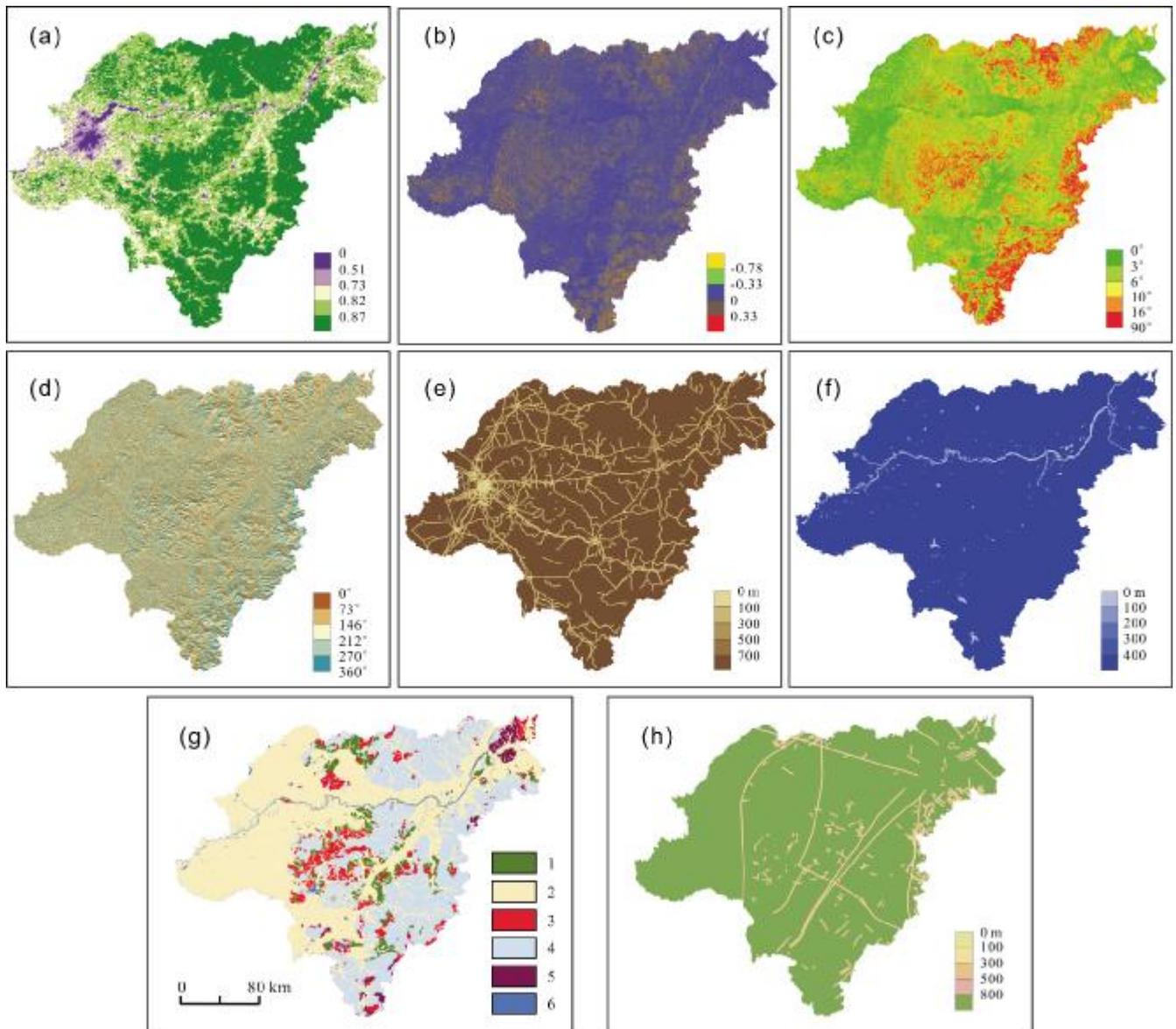


图 2 研究区评价因子分层图

Fig. 2 Evaluation factor layers of the study area

a—归一化植被指数(NDVI); b—曲率(curvature); c—坡度(slope gradient); d—坡向(slope aspect); e—距道路距离(distance from road); f—距水系距离(distance from river); g—地层(strata); h—距构造距离(distance from structure); 1—变质岩类(metamorphic rock); 2—侵入岩类(intrusive rock); 3—喷出岩类(effusive rock); 4—第四系(Quaternary); 5—沉积岩类(sedimentary rock); 6—水系(water system)

样本. 将负样本与正样本组合建立论文数据集, 其中正样本记为 1, 负样本记为 0, 数据集空间分布如图 3a 所示. 随机选取数据集中 70% 作为训练集, 30% 作为测试集.

4.2 实验环境设置

本研究使用 Python 语言的 Sklearn 库进行算法构建, 采用学习曲线和网格搜索进行参数调优. 各算法参数设置如下: 随机森林算法中 $n_estimators=144$, $max_depth=10$, $max_features=7$; Adaboost 算法中

$base_estimator$ 为决策树, $max_depth=3$, $n_estimators=81$; Stacking 算法中选择随机森林、支持向量机和 Adaboost 算法为基学习器, 选择逻辑回归算法为元学习器.

4.3 地质灾害易发性评价

利用集成学习算法对研究区地质灾害易发性进行预测, 根据地质灾害易发概率, 利用 ArcGIS 软件以 0.2 为步长, 进行地质灾害易发性分区图制作. 将地质灾害易发性分区图划分为极低易发区、低易发区、中易发

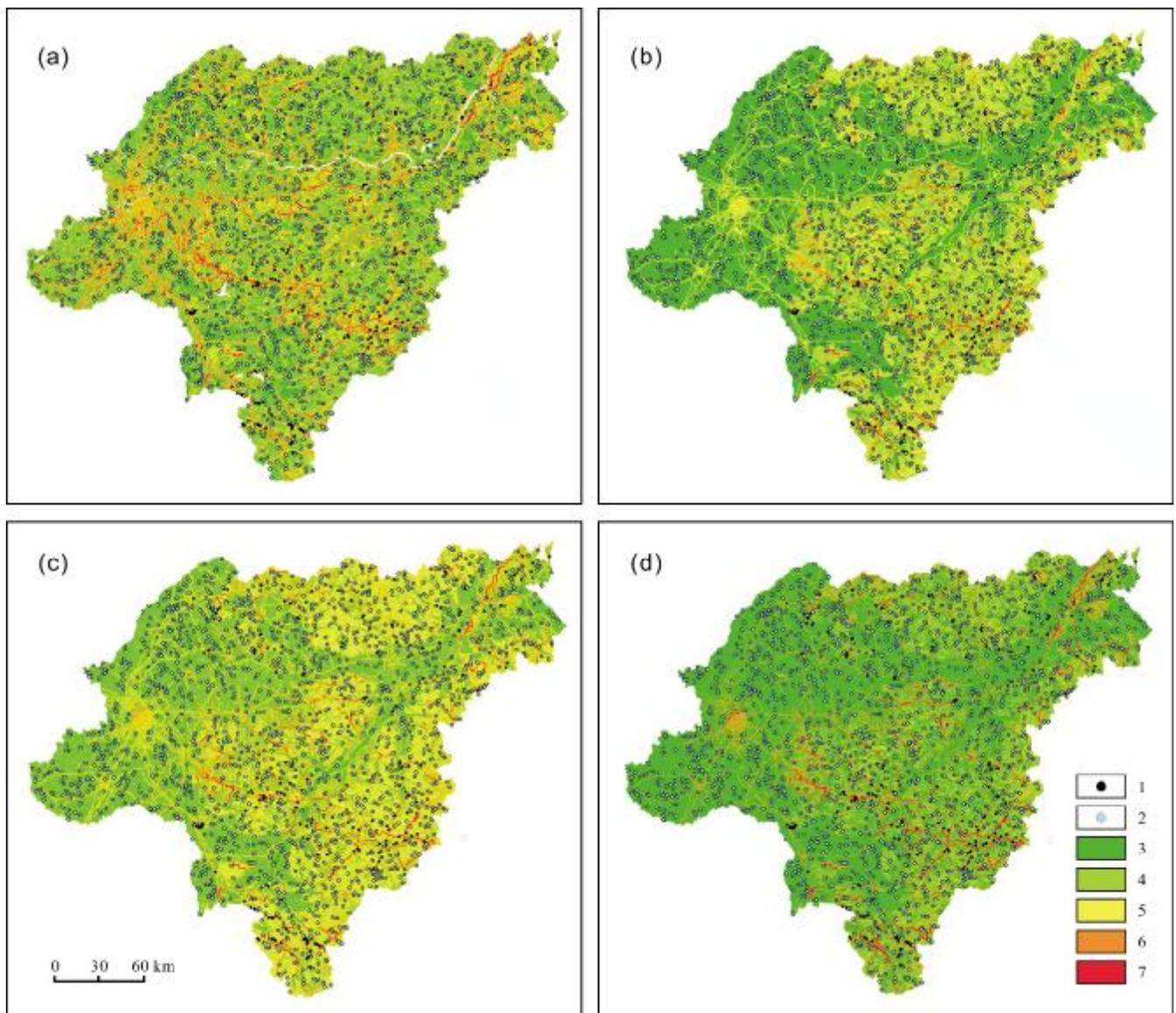


图 3 各评价方法预测结果图

Fig. 3 Prediction results of different evaluation methods

a—信息量算法(information algorithm); b—随机森林算法(random forest algorithm); c— Adaboost 算法(Adaboost algorithm); d— Stacking 算法(Stacking algorithm); 1—地质灾害点 (geohazard site); 2—非地质灾害点 (non-geohazard site); 3—极低易发区 (extremely low susceptible zone); 4—低易发区 (low susceptible zone); 5—中等易发区 (medium susceptible zone); 6—高易发区 (high susceptible zone); 7—极高易发区 (extremely high susceptible zone)

区、高易发区、极高易发区 5 类,结果如图 3b—d.

由图 3 可知,信息量算法和集成学习算法对研究区的地质灾害易发区评价结果具有较高的相似性.结合研究区评价因子分层图可得出极高易发区和高易发区主要沿道路展布,与实际情况相吻合.哈尔滨市地质灾害主要诱发因素之一是开山修路导致的山体结构变形,为地质灾害发生提供了基础.

将已知地质灾害点数据在各个分区中的占比进行空间统计分析,结果见表 3. 为了更直观地展示各算法的分区预测结果精度,采用频率比^[29]进行评价. 频率比(FR)在算法预测精度评价中具有广泛的适用性. $FR \geq 1$,表示分区与地质灾害的相关性高; $FR < 1$,表示

与地质灾害的相关性低. 计算公式如下:

$$FR_i = \frac{M_i/M}{N_i/N} \quad (2)$$

式中, M_i 是 i 类易发性等级中地质灾害发生数量, M 为研究区总的地质灾害发生数量, N_i 是 i 类易发性等级中栅格数量, N 为研究区总的栅格数量.

由表 3 可知,各算法随着危险等级的提升,地质灾害与评价等级面积占比的频率比均在提升,表明 4 种算法均能对预测区地质灾害易发性进行较好的预测. 信息量算法的极高易发区面积占比 3.04%, 发生地质灾害 63 个,频率比 9.67; 极低易发区面积占比 31.92%, 发生地质灾害 15 个,频率比 0.22. 随机森林算法的极

表 3 基于不同算法的易发性分区结果

Table 3 Susceptibility zoning results based on different algorithms

评估算法	危险等级	栅格单元数据量	栅格单元面积比例/%	地质灾害数量	地质灾害数量占比/%	频率比
信息量	极低易发	18604798	31.92	15	7.01	0.22
	低易发	21629659	37.11	37	17.29	0.47
	中易发	11344966	19.46	53	24.77	1.27
	高易发	4931899	8.46	46	21.50	2.54
	极高易发	1773834	3.04	63	29.44	9.67
随机森林	极低易发	2434647	45.75	31	14.49	0.32
	低易发	1378756	25.91	33	15.42	0.60
	中易发	1102540	20.72	45	21.03	1.02
	高易发	338744	6.36	60	28.04	4.41
	极高易发	67501	1.27	45	21.03	16.58
Adaboost	极低易发	446386	8.39	4	1.87	0.22
	低易发	2493305	46.85	49	22.90	0.49
	中易发	1797451	33.77	72	33.64	1.00
	高易发	515218	9.68	55	25.70	2.65
	极高易发	69828	1.31	34	15.89	12.11
Stacking	极低易发	3171461	59.59	45	21.03	0.35
	低易发	663386	12.46	23	10.75	0.86
	中易发	319539	6.00	13	6.07	1.01
	高易发	421394	7.92	24	11.21	1.42
	极高易发	746408	14.02	109	50.93	3.63

高易发区面积占比 1.27%,发生地质灾害 45 个,频率比 16.58;极低易发区面积占比 45.75%,发生地质灾害 31 个,频率比 0.32. Adaboost 算法的极高易发区面积占比 1.31%,发生地质灾害 34 个,频率比 12.11;极低易发区面积占比 8.39%,发生地质灾害 4 个,频率比 0.22. Stacking 算法的极高易发区面积占比 14.02%,发生地质灾害 109 个,频率比 3.63;极低易发区面积占比 59.59%,发生地质灾害 45 个,频率比 0.35.

尽管信息量算法和 Stacking 预测出的极高易发区的地质灾害数量高于随机森林算法,但所占面积比随机森林算法更高. 随机森林算法的频率比明显高于其他算法,即使用更少的预测区域尽可能地预测了更多的地质灾害发生情况. 因此,预测精度的结果排序为随机森林算法>Adaboost 算法>信息量算法>Stacking 算法.

4.4 算法准确性指标

算法准确性评价对于地质灾害易发性评价结果至关重要. 混淆矩阵^[30]常用于分类问题的预测精度分析,由该矩阵得到的接收灵敏度(ROC)曲线、AUC、准确率、精度、F1 分数作为评估分类结果的重要指标. ROC 曲线的横轴代表非灾害点预测为灾害点的比例,纵坐标表示灾害点预测准确性的占比,ROC 曲线越靠近 Y 坐标轴,说明算法的预测性能越优;曲线下方的面积(即 AUC)代表算法的准确性,AUC 的取值范围为 0~1,值越靠近 1,表现算法性能越优(图 4).

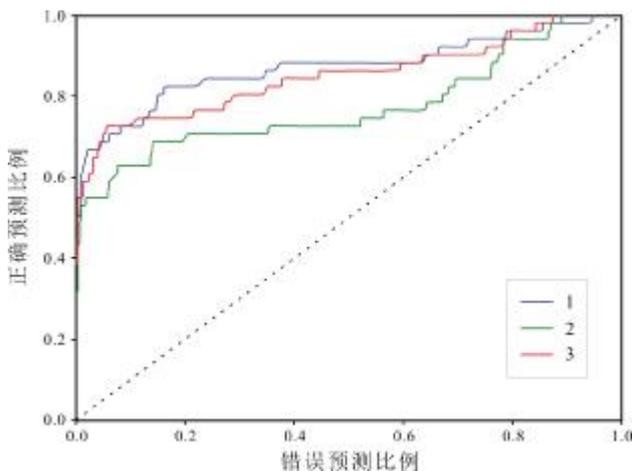


图 4 不同算法 ROC 曲线评价结果

Fig. 4 ROC curves of evaluation results of different algorithms

1—随机森林模型(random forest model); 2—Adaboost 模型(Adaboost model); 3—Stacking 模型(Stacking model)

利用数据集中的测试集对各算法的指标进行计算,评价结果如表 4. 由表 4 可知,Stacking 算法在测试集中的表现除 AUC 指标外均略优于随机森林算法,明显好于 Adaboost 算法. 结合表 4 以及 Stacking 算法原理分析,推测 Stacking 集合过程中对基学习器(随机森林算法、Adaboost 算法、SVC 算法)的预测结果继续进行特征提取,导致其更多地识别出地形因子对地质灾害易发性评估的影响,因此在测试集中 Stacking 算法的准确性指标表现更好,也是此原因造成 AUC 值略低于随机森林 AUC 值. 但在区域预测时,将更多的地形线密集区划分为极高风险区,导致极高易发区面积占比达 14.02%,表明 Stacking 对训练集数据的识别存在一定的过拟合现象.

表 4 算法预测性能结果

Table 4 Prediction performance results of different algorithms

算法	准确率	精度	AUC	F1
随机森林	0.941	0.893	0.891	0.633
Adaboost	0.923	0.683	0.742	0.609
Stacking	0.953	0.938	0.855	0.723

5 结论

1)以黑龙江省哈尔滨市为研究区,基于因子相关性分析,选择坡度、坡向、曲率、岩性、NDVI、距水系距离、距道路距离、距构造距离等 8 类评价因子建立易发性评价指标体系,利用信息量算法、随机森林算法、Adaboost 算法、Stacking 算法开展研究区地质灾害易发性评价,结果表明 4 种算法易发性评价分区图与研究区实际情况较为一致. 哈尔滨市地质灾害主要诱发因素为人类工程活动,极高发区主要集中在道路附近.

2)在数据集建立过程中,依据信息量算法评价结果中各易发性等级比例,在极低易发区和低易发区随机选取间隔大于 1 km 的非地质灾害点,可有效避免因数据冗余造成的算法分类精度下降.

3)通过评价结果和准确性指标验证,集成学习算法能够应用于地质灾害易发性预测中,但应注意过拟合风险. 随机森林算法更适用于本区域的地质灾害易发性评价.

参考文献(References):

- [1] 杨锐, 于春勇, 汪洋. 基于加权信息量和 GIS 方法的地质灾害易发性评价——以榆社县为例[J]. 矿业研究与开发, 2023, 43(2): 163-170.
Yang R, Yu C Y, Wang Y. Susceptibility evaluation of geological disaster based on weighted information and GIS method: A case study of Yushu County[J]. Mining Research and Development, 2023, 43(2): 163-170.
- [2] 窦杰, 向子林, 许强, 等. 机器学习在滑坡智能防灾减灾中的应用与发展趋势[J]. 地球科学, 2023, 48(5): 1657-1674.
Dou J, Xiang Z L, Xu Q, et al. Application and development trend of machine learning in landslide intelligent disaster prevention and mitigation[J]. Earth Science, 2023, 48(5): 1657-1674.
- [3] 莫运松, 江思义, 邹仁辉, 等. 基于专家-层次分析法的地质灾害危险性评价——以广西富川瑶族自治县为例[J]. 矿产勘查, 2023, 14(2): 293-303.
Mo Y S, Jiang S Y, Zou R H, et al. Risk assessment of geological disasters based on expert-analytic hierarchy process: A case study of Fuchuan Yao Autonomous County in Guangxi[J]. Mineral Exploration, 2023, 14(2): 293-303.
- [4] Zhu X G, Yan R. Risk assessment of geological hazards in the Dahe Mountain area based on AHP[J]. DEStech Transactions on Environment Energy and Earth Science, 2019. DOI: 10.12783/dteees/gmce2018/27463.
- [5] 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, 340.
Mao Z J, Zhang J G, Zhong J X, 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, 340.
- [6] Ma J R, Wang X D, Yuan G X. Evaluation of geological hazard susceptibility based on the regional division information value method [J]. ISPRS International Journal of Geo-Information, 2023, 12(1): 17.
- [7] 宋渊, 江南, 张恩博, 等. 基于证据权法的湖北省兴山县滑坡灾害易发性评价[J]. 地质与资源, 2023, 32(3): 352-359, 289. DOI: 10.13686/j.cnki.dzyzy.2023.03.012.
Song Y, Jiang N, Zhang E B, et al. Susceptibility evaluation of landslide hazard in Xingshan County of Hubei Province based on weights-of-evidence method [J]. Geology and Resources, 2023, 32(3): 352-359, 289. DOI: 10.13686/j.cnki.dzyzy.2023.03.012.
- [8] Agrawal N, Dixit J. GIS-based landslide susceptibility mapping of the Meghalaya-Shillong Plateau region using machine learning algorithms [J]. Bulletin of Engineering Geology and the Environment, 2023, 82(5): 170.
- [9] 张文龙, 张振凯, 杨帅. 勉略宁地区地质灾害危险性智能评价和区划研究[J]. 西北地质, 2023, 56(1): 276-283.
Zhang W L, Zhang Z K, Yang S. Study on intelligent evaluation and zoning of geohazards risk in Mianluening area [J]. Northwestern Geology, 2023, 56(1): 276-283.
- [10] 刘海知, 徐辉, 包红军, 等. 机器学习分类算法在降雨型滑坡预报中的应用[J]. 应用气象学报, 2022, 33(3): 282-292.
Liu H Z, Xu H, Bao H J, et al. Application of machine learning classification algorithm to precipitation-induced landslides forecasting [J]. Journal of Applied Meteorological Science, 2022, 33(3): 282-292.
- [11] Wang J Q, Sun P F, Chen L L, et al. Recent advances of deep learning in geological hazard forecasting [J]. Computer Modeling in Engineering & Sciences, 2023, 137(2): 1381-1418.
- [12] Fang Z C, Wang Y, Peng L, et al. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping [J]. International Journal of Geographical Information Science, 2021, 35(2): 321-347.
- [13] 茹曼, 郑燕, 张斌, 等. 基于 SVM-RF 模型的地质灾害易发性评价——以河南省博爱县青天河景区为例[J]. 地质与资源, 2023, 32(5): 633-641. DOI: 10.13686/j.cnki.dzyzy.2023.05.014.
Ru M, Zheng Y, Zhang B, et al. Assessment of geological disaster susceptibility based on SVM-RF model: A case study of Qingtian River scenic area in Boai County, Henan Province [J]. Geology and Resources, 2023, 32(5): 633-641. DOI: 10.13686/j.cnki.dzyzy.2023.05.014.
- [14] 徐继伟, 杨云. 集成学习方法: 研究综述[J]. 云南大学学报(自然科学版), 2018, 40(6): 1082-1092.
Xu J W, Yang Y. A survey of ensemble learning approaches [J]. Journal of Yunnan University, 2018, 40(6): 1082-1092.
- [15] 虞瑶, 高涵, 陶畅, 等. 基于集成学习的高光谱遥感影像分类综述 [J]. 测绘与空间地理信息, 2023, 46(4): 49-52, 60.
Yu Y, Gao H, Tao Y, et al. A review of hyperspectral remote sensing image classification based on ensemble learning [J]. Geomatics & Spatial Information Technology, 2023, 46(4): 49-52, 60.
- [16] 苏彬. 黑龙江省哈尔滨市市区地质灾害研究[D]. 长春: 吉林大学, 2014.
Su B. Research on geological hazard in Harbin city of Heilongjiang province [D]. Changchun: Jilin University, 2014.
- [17] 王佳佳, 殷坤龙, 肖莉丽. 基于 GIS 和信息量的滑坡灾害易发性评价——以三峡库区万州区为例[J]. 岩石力学与工程学报, 2014, 33(4): 797-808.
Wang J J, Yin K L, Xiao L L. Landslide susceptibility assessment based on GIS and weighted information value: A case study of Wanzhou District, Three Gorges reservoir [J]. Chinese Journal of Rock Mechanics and Engineering, 2014, 33(4): 797-808.
- [18] Ling X, Zhu Y Q, Ming D P, et al. Feature engineering of geohazard susceptibility analysis based on the random forest algorithm: Taking Tianshui City, Gansu Province, as an example [J]. Remote Sensing,

- 2022, 14(22): 5658.
- [19] Ramineni V, Surekha Y, Kumar A V. Machine learning based Adboost algorithms [J]. *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, 7(6S5): 1928–1932.
- [20] 郭亚雷, 邓念东, 李宇新, 等. 基于堆叠模型的滑坡易发性评价——以商洛市丹凤县为例 [J]. *自然灾害学报*, 2023, 32(2): 243–252.
- Guo Y L, Deng N D, Li Y X, et al. Evaluation of landslide susceptibility based on Stacking: Take Danfeng County, Shangluo City as an example [J]. *Journal of Natural Disasters*, 2023, 32(2): 243–252.
- [21] Ting K M, Witten I H. Issues in stacked generalization [J]. *Journal of Artificial Intelligence Research*, 1999, 10(1): 271–289.
- [22] Shirzadi A, Bui D T, Pham B T, et al. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach [J]. *Environmental Earth Sciences*, 2017, 76(2): 60.
- [23] 周侯伯, 肖桂荣, 林炫歆, 等. 基于特征筛选与差分进化算法优化的滑坡危险性评估方法 [J]. *地球信息科学学报*, 2022, 24(12): 2373–2388.
- Zhou H B, Xiao G R, Lin X X, et al. Landslide hazard assessment method based on feature screening and differential evolution algorithm optimization [J]. *Journal of Geo-information Science*, 2022, 24(12): 2373–2388.
- [24] 孔嘉旭, 庄建琦, 彭建兵, 等. 基于信息量和卷积神经网络的黄土高原滑坡易发性评价 [J]. *地球科学*, 2023, 48(5): 1711–1729.
- Kong J X, Zhuang J Q, Peng J B, et al. Evaluation of landslide susceptibility in Chinese Loess Plateau based on IV-RF and IV-CNN coupling models [J]. *Earth Science*, 2023, 48(5): 1711–1729.
- [25] 戴梦兰, 陈志波, 施国栋. 基于分形理论的浙江省地质灾害分布特征与影响因素分析 [J]. *中国地质灾害与防治学报*, 2022, 33(6): 63–69.
- Dai M L, Chen Z B, Shi G D. Analysis of the distribution characteristics and influencing factors of geological hazards in Zhejiang Province based on fractal theory [J]. *The Chinese Journal of Geological Hazard and Control*, 2022, 33(6): 63–69.
- [26] Wu X Y, Song Y B, Chen W, et al. Analysis of geological hazard susceptibility of landslides in Muli County based on random forest algorithm [J]. *Sustainability*, 2023, 15(5): 4328.
- [27] 左群超, 叶天竺, 冯艳芳, 等. 中国陆域 1:25 万分幅建造构造图空间数据库 (V1) [EB/OL]. 中国地质调查局发展研究中心 [创建机构], 2006. <http://dcc.ngac.org.cn/geologicalData/rest/geologicalData/geologicalDataDetail/8adaeff963f2eb2a0163f335b7600019>.
- Zuo Q C, Ye T Z, Feng Y F, et al. Spatial data base of serial suite-tectonic map-sheets of mainland China (1 : 250 000) [EB/OL]. *Geology in China*. 2018. 45(S1): 1–26. (in Chinese)
- [28] 董金玮, 周岩, 尤南山, 等. 2000—2022 年中国 30 米年最大 NDVI 数据集 [DB/OL]. 国家生态科学数据中心, 2021. <https://doi.org/10.12199/nesdc.ecodb.rs.2021.012>.
- Dong J W, Zhou Y, You N S, et al. Divergent shifts in peak photosynthesis timing of temperate and alpine grasslands in China [DB/OL]. *Remote Sensing of Environment*, 2021. <https://www.nesdc.org.cn/sdo/detail?id=60f68d757e28174f0e7d8d49>. (in Chinese)
- [29] Shahabi H, Hashim M, Ahmad B B. Correction to: Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran [J]. *Environmental Earth Sciences*, 2023, 82(11): 281.
- [30] 杨硕, 李德营, 严亮轩, 等. 基于随机森林模型的乌江高陡岸坡滑坡地质灾害易发性评价 [J]. *安全与环境工程*, 2021, 28(4): 131–138.
- Yang S, Li D Y, Yan L X, et al. Landslide susceptibility assessment in high and steep bank slopes along Wujiang River based on random forest model [J]. *Safety and Environmental Engineering*, 2021, 28(4): 131–138.