

基于 WOE-CatBoost 耦合模型的滑坡灾害 易发性评价

刘娟, 王宇栋, 李晓丽, 林静

昆明理工大学公共安全与应急管理学院, 昆明 650093

摘要: 宣威市地形地貌复杂, 岩质较弱, 易发生滑坡。本文以宣威市为例, 选取地形地貌、气候水文和地质构造等 12 个指标, 采用证据权模型(WOE), 得到各指标分级状态的证据权重; 再通过 WOE 模型与 CatBoost 算法耦合, 得到研究区滑坡灾害易发性评价指数; 最后选取 SHAP 对滑坡影响因子进行全局和局部解释。研究结果显示: (1) 采用 WOE 和 WOE-CatBoost 耦合模型 AUC 值分别 0.845 5 和 0.984 6, 滑坡极高易发区的滑坡密度分别为 0.181 0 个/km² 和 0.297 0 个/km²; (2) 高程、地层岩性、NDVI 和降雨是影响滑坡发生的主要因素。得到结论 WOE-CatBoost 模型可以用于滑坡易发性的预测, 且滑坡极高、高易发区主要分布于研究区的东南部和东北部, 低易发区主要分布在研究区的中部和西南部。

关键词: 滑坡灾害; 机器学习; SHAP; 易发性; 预测精度; WOE

中图分类号: P642.22 **文献标识码:** A

Landslide Hazard Susceptibility Assessment Based on A Coupled WOE-CatBoost Model

LIU Juan, WANG Yudong, LI Xiaoli, LIN Jing

School of Public Safety and Emergency Management, Kunming University of Science and
Technology, Kunming 650093, China

Abstract: Xuanwei City is characterized by complex topography and geomorphology, weak lithology, and frequent landslide occurrences. This study selects 12 evaluation factors, including topography, climatic hydrology, and geological structure in Xuanwei City, and applies the Weight of Evidence (WOE) model to obtain the evidential weights of each factor's graded states. A coupled WOE-CatBoost model is subsequently developed to derive landslide susceptibility indices for the study area. SHAP is further employed for global and local interpretability analyses of landslide-influencing factors. Results demonstrate: (1) The WOE and WOE-CatBoost models achieve AUC values of 0.845 5 and 0.984 6, respectively, with landslide densities in extremely high susceptibility zones reaching 0.181 0/km² and 0.297 0/km²; (2) Elevation, lithology, NDVI, and rainfall are identified as primary controlling factors for landslide occurrence. The findings confirm that the WOE-CatBoost model effectively predicts landslide susceptibility, with extremely high and high

收稿日期: 2024-07-19 改回日期: 2024-09-03

基金项目: 云南省科技厅基础研究计划项目(202401CF070138)

第一作者简介: 刘娟(1997-), 女, 硕士研究生, 研究方向: 应急管理与风险分析。E-mail: 157020636@qq.com

引文格式: 刘娟, 王宇栋, 李晓丽, 等. 基于 WOE-CatBoost 耦合模型的滑坡灾害易发性评价[J]. 地质灾害与环境保护, 2025, 36(1): 6-14.

LIU Juan, WANG Yudong, LI Xiaoli, et al. Landslide hazard susceptibility assessment based on a coupled WOE-CatBoost model[J].
Journal of Geological Hazards and Environment Preservation, 2025, 36(1): 6-14.

susceptibility zones predominantly distributed in southeastern and northeastern regions, while low susceptibility areas are concentrated in central and southwestern regions.

Key words: Landslide disaster; Machine learning; SHAP; Susceptibility; Prediction accuracy; WOE

0 引言

我国是全世界滑坡灾害发生最频繁的地方,造成的财产损失不计其数,人们的生命安全也受到威胁^[1]。为降低滑坡灾害造成的损失,需要对滑坡灾害区域进行易发性评估,以便进行更好的防治,达到以人为本的目的。

滑坡灾害易发性评价方法主要分为定性和定量评价方法。定性分析方法如朱晓曦^[2]选取信息量模型结合专家系统改进后得到文成县滑坡灾害综合指标计算模型。Adnan Ozdemir^[3]对频率比、证据权重和逻辑回归生成的滑坡敏感性图进行了评估和比较,研究结果显示频率比与证据权重之间存在高度相关性。定量分析方法多采用基于机器学习的统计模型。夏辉^[4]选用向量机(SVM)和人工神经网络(ANN)模型,得到巫山县滑坡易发性区划图。研究表明,SVM模型相比于ANN模型,在成功率和预测率曲线的AUC值方面表现更优。林报嘉^[5]尝试将新型的XGBoost机器学习模型应用于滑坡灾害易发性研究,结果表明,XGBoost模型的训练集和验证集的ROC曲线下面积均优于SVM模型。Hussain^[6]将RF、XGBoost和KNN应用于中巴喀喇昆仑公路沿线,其对预测滑坡灾害有着良好的效果。Zhang^[7]利用最新的LightGBM和CatBoost等算法,在万州地区滑坡易发性评估中实现了更高的精度,有效地解决了分类特征、梯度偏差和预测偏移等相关问题,提高了算法的精度和泛化能力。然而,随着机器学习模型复杂度的增加,其预测结果往往表现出一定的“黑盒”特性,缺乏可解释性,限制了其在实际应用中的可信度和接受度。传统模型如证据权只是将环境因子转换为反映其对滑坡影响的权重值,从而衡量类别变量对滑坡预测的影响程度,却不能直接具体到某一个数值。本研究引入了SHAP(SHapley Additive exPlanations)方法,对模型的预测结果进行解释。相比传统的特征重要性分析方法,SHAP通过计算每个特征值对模型输出的贡献,提供了特征的重要性排序,并定量地揭示每个特征对预测结果的具体贡献。这不仅揭示了模型预测的原因,还明确了特征之间的关系,提供了更为可信的滑坡易发性评估结果。

综上所述,本文以宣威市作为研究区域,通过收集相关的地形地貌、地质构造、气候水文等数据资料,结合云南宣威市的特点,选取了12个评价因子构建了滑坡易发性评价指标体系。数据处理后,利用WOE-CatBoost模型进行滑坡模型的构建和评价因子重要性分析。

1 研究区概况及数据来源

1.1 研究区概况

研究区位于云贵高原中部,地理坐标东经 $103^{\circ}35'30''\sim 104^{\circ}40'50''$,北纬 $25^{\circ}53'30''\sim 26^{\circ}44'50''$ 之间,属于温带季风气候,年平均降水量约为940 mm。地势则西部由东北向西南逐渐下降,东部由西南向东北逐渐下降(图1)。岩体软硬相间,整体岩质相对较软弱,自身抗风化能力较弱,表层岩体受物理风化作用强烈影响,力学性质较差,工程地质性质也较差,加上不合理的人类工程活动,地质灾害相对频发。通过揭示滑坡灾害危险分布规律,可有效减少人员伤亡风险,提升防灾和应对能力,保障地区居民和社会安全。

1.2 数据来源

研究数据包括遥感影像数据、滑坡灾害隐患点数据和河流路网数据等。坡度、坡向数据是通过数字高程模型(DEM)数据提取处理得到的,栅格数据采用30 m分辨率,WGS_1984_UTM_Zone_50N的投影坐标系统。具体数据来源请参考表1。

表1 研究区数据来源

评价因子	数据类型	数据来源
高程	栅格	地理空间数据云
滑坡灾害点	矢量	云南省自然资源厅
降雨	文本	宣威市气象局
地层岩性、断层	矢量	中国地质调查云
植被、土地利用、土壤侵蚀、土壤类型	栅格	国家生态数据科学中心
道路分布图	矢量	OpenStreetMap

2 研究方法

2.1 证据权法

证据权法是一种离散的多元统计方法,最初用于医疗诊断支持^[8]。该方法基于贝叶斯概率统计原理,通过计算各因素与滑坡灾害发生的条件概率,评

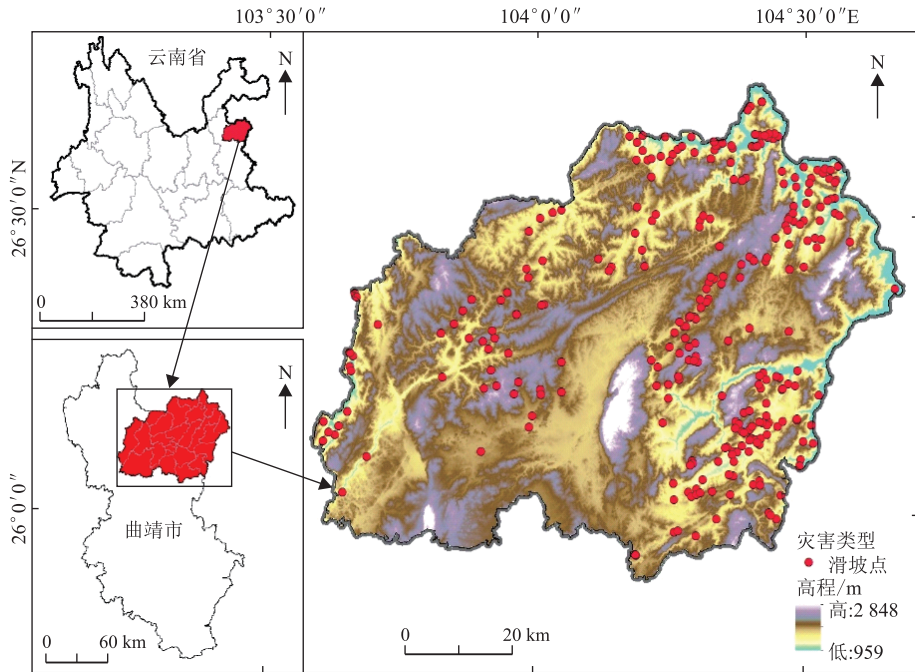


图1 研究区地理位置及滑坡点分布

估各因素对滑坡灾害发生的贡献程度。其计算流程如下:

$$W_+ = \ln \frac{P(B/D)}{P(\bar{B}/D)} \quad (1)$$

$$W_- = \ln \frac{P(\bar{B}/D)}{P(B/D)} \quad (2)$$

$$C = W_+ - W_- \quad (3)$$

式中, P 代表滑坡事件发生的概率; B 表示具有滑坡因素的区域单元数; D 表示已经发生滑坡的单元数; W_+ 表示正相关权重, W_- 表示负相关权重。通过计算正负相关权重之间的差值, 即对比度 C , 评估证据因子与数据集之间的相关性。证据权重的值越大, 说明该分级状态与滑坡的关联程度越高。

2.2 CatBoost

CatBoost (Categorical Boosting) 是一种梯度提升算法, 用于处理具有大量类别型特征的数据集^[9]。与传统的梯度提升算法相比, CatBoost 能够自动处理类别型特征^[10], 无需进行独热编码, 具有高效性能、防止过拟合的特性以及快速训练的优势。通过基于梯度提升算法的优化技术和对称树处理类别特征的方法, CatBoost 在分类问题中表现出色, 提供了高准确性和可解释性。

2.3 SHAP

SHAP 方法源自博弈论中的 Shapley 值概念^[11], 用于解释机器学习模型预测。在面对复杂的黑盒模型时, SHAP 方法通过计算每个特征值对于模型输出的贡献, 为每个特征生成相对重要性的解

释^[12], 帮助理解模型预测的原因和特征之间的关系。

3 评价因子选取及独立性检验

在地质灾害易发性评价研究中, 常见的单元划分方法包括斜坡单元、行政单元和栅格单元等。本文选择像元大小为 $30 \text{ m} \times 30 \text{ m}$ 的栅格单元作为评价单元, 通过矢量转点得到了 6 959 485 个栅格单元用于进行滑坡灾害的易发性评估和分析。这种栅格单元的划分能够提供相对均匀且具有一定空间范围的单元, 从而确保评价结果的准确性和可比性。

3.1 评价因子选取与处理

评价指标因子的选取对研究区滑坡灾害的危险区的划分有着重要影响^[13, 14]。综合考虑研究区的地质、地形和气候等多方面因素, 选择了以下高程、坡度、坡向、地层岩性等 12 个与滑坡灾害相关的评价因子(图 2)。

研究区域内地形地貌主要包括高程、坡度、坡向, 数据均为连续性数据。当断层出现错动和滑动时, 土体结构遭到破坏, 稳定性降低, 削弱了土壤的抗剪强度, 使得断层附近的土体更容易发生滑动。不同的地层岩性具有不同的力学特性和稳定性, 直接影响着滑坡的形成和演化, 将研究区分为松散土体岩组、碎屑岩类、碳酸盐类和岩浆岩类 4 种二级状态。滑坡在碎屑岩和岩浆岩类最为发育。植物的根

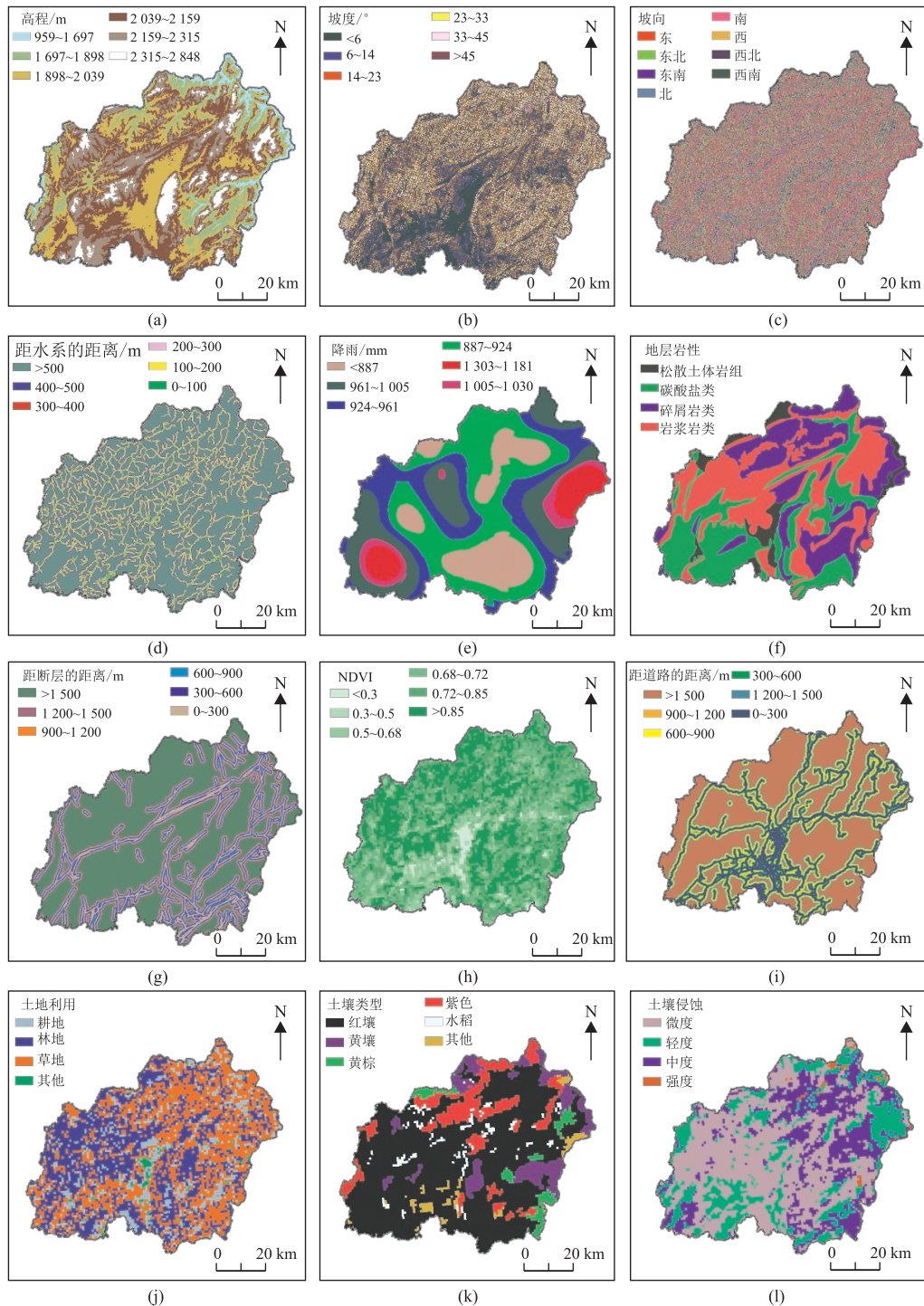


图 2 滑坡易发性评价因子分级

系通过穿透和纠结在土壤中,形成一个复杂的根网结构,有效地抵抗外部力量对土壤的剪切作用。不同的土壤类型对滑坡发生具有重要影响。黏土质土壤通常具有较好的保水性和抗侵蚀能力,而砂质土壤则易发生坍塌。水流对土壤的侵蚀作用使土壤表层逐渐流失,导致土壤结构疏松,抗剪强度降低,从而削弱了土壤的稳定性,增加了土地滑坡的风险。研究区主要受水力侵蚀的影响,按侵蚀程度分为微

度、轻度、中度、强度。土地的过度开垦和不合理的利用方式加剧了土壤侵蚀的程度,导致土地失去稳定性,增加了土地的脆弱性,进而促使滑坡的发生。其中,地层岩性、土地利用、土壤类型和土壤侵蚀均为非连续型影响因子。

3.2 评价因子独立性检验

利用 SPSS 软件对滑坡点获取的指标因子进行了皮尔逊相关性分析,具体结果见表 2。当相关系

数 $|R|$ 超过 0.3 时,可认为因子间具有较强的相关性^[15]。研究结果显示,所有评价因子之间的相关性均小于 0.3,表明所选取的评价因子间关联较弱,适合用于进一步的评估和分析。

表 2 评价因子相关性分析

评价因子	高程/m	坡度/°	坡向	距水系的距离/m	降雨/mm	地层岩性	距断层的距离/m	NDVI	距道路的距离/m	土地利用	土壤类型	土壤侵蚀
高程/m	1.00	-0.04	-0.07	-0.09	-0.18	0.19	-0.12	-0.04	-0.06	0.10	-0.24	-0.11
坡度/°		1.00	0.18	-0.04	0.07	-0.04	0.03	0.12	-0.09	0.01	0.08	-0.06
坡向			1.00	0.03	-0.02	-0.04	0.05	-0.01	0.04	0.02	0.03	0.00
距水系的距离/m				1.00	-0.08	0.03	0.03	-0.11	0.07	-0.04	-0.09	0.01
降雨/mm					1.00	0.09	-0.01	-0.04	-0.10	-0.09	0.08	0.00
地层岩性						1.00	-0.10	-0.12	0.00	-0.09	-0.20	-0.03
距断层的距离/m							1.00	-0.07	0.15	-0.02	0.05	0.10
NDVI(植被归一化指数)								1.00	-0.25	-0.06	0.12	-0.08
距道路的距离/m									1.00	-0.01	-0.04	0.04
土地利用										1.00	0.05	0.14
土壤类型											1.00	0.15
土壤侵蚀												1.00

4 滑坡灾害易发性评价与结果分析

4.1 滑坡易发性建模

在研究区域内,利用 ArcGIS 按照滑坡点与非滑坡点比值为 1 : 2 的比例随机选取非滑坡点。将滑坡点标记为“0”,非滑坡点标记为“1”,共构建了

852 个数据样本。按照公式(3)算出各因子的证据权重值,见表 3。将得到的证据权重值使用空间连接赋值于各指标因子二级状态的栅格点中,使用 ArcGIS 分析工具中的栅格计算器将各因子叠加分析,得到研究区滑坡灾害的总证据权重值。应用自然间断点法将研究区易发性分为 5 个等级得到滑坡预测的 WOE 模型。

表 3 评价因子二级状态证据权重值

评价因子	分级	证据权重值	评价因子	分级	证据权重值	评价因子	分级	证据权重值
高程/m	<1 697	1.553	降雨/mm	<887	-0.657	地层岩性	松散土体岩	-2.859
	1 697~1 898	1.933		887~924	0.049		碎屑岩类	1.022
	1 898~2 039	-0.143		924~961	0.659		碳酸盐类	-1.626
	2 039~2 159	-0.826		961~1 005	0.296		岩浆岩类	0.484
	2 159~2 315	-1.987		1 005~1 030	-0.495		耕地	0.630
>2 315	-2.226	1 030~1 181	-2.465	土地利用	林地	-0.784		
坡度/°	<6	-0.491	距断层的距离/m		0~300	0.212	草地	0.307
	6~14	-0.616			300~600	0.236	其他	-0.661
	14~23	0.436			600~900	0.469	红壤	-1.227
	23~33	0.680			900~1 200	0.305	黄壤	0.770
	33~45	-0.277		1 200~1 500	-0.049	黄棕壤	-0.616	
>45	0.571	>1 500	-0.622	土壤类型	紫色土	1.039		
坡向	东北	-0.270	NDVI		<0.3	-0.620	水稻土	0.012
	东	-1.038			0.3~0.5	-0.848	其他	-0.446
	东南	-0.811			0.5~0.68	-0.531	微度	-0.672
	南	-0.930			0.68~0.73	0.172	轻度	0.389
	西南	-1.321		0.72~0.85	0.630	中度	0.117	
	西	-1.081	>0.85	-0.438	强度	1.660		
	西北	-1.244	距道路的距离/m	0~300	0.244			
	北	-0.325		300~600	0.040			
距水系的距离/m	0~100	0.285		600~900	0.060			
	100~200	0.891		900~1 200	-0.418			
	200~300	0.190		1 200~1 500	-0.120			
	300~400	0.282	>1500	-0.037				
	400~500	-0.610						
>500	-0.730							

WOE-CatBoost 模型是以各因子得到的证据权重值作为输入,按照 8 : 2 比例将数据分为训练集和

测试集。使用 SMOTE 对数据样本进行不平衡处理,利用网格交叉进行调整参数范围。将研究区的

栅格提取各评价因子的属性值到点,生成表格,导入训练好的 CatBoost 模型中,预测每个栅格点发生滑坡的概率,利用点转栅格工具将所有的点生成栅格

数据,再用自然间断法将研究区内的滑坡易发区分为极高、高、中、低和极低易发区 5 个类别。

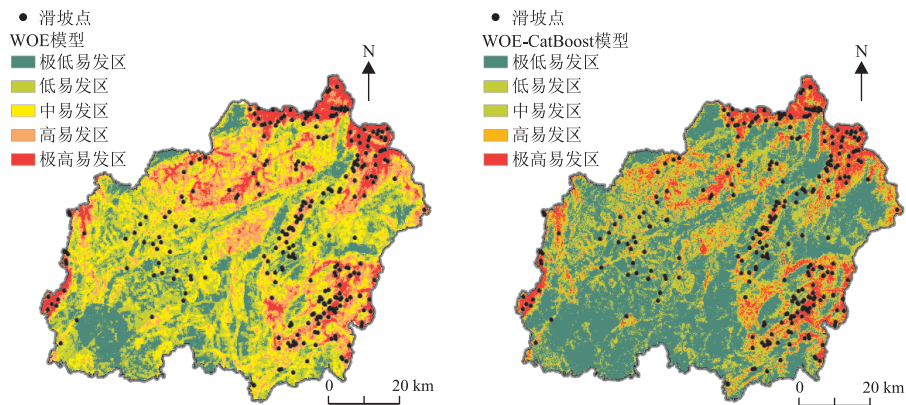


图 3 滑坡易发性分区图

4.2 模型性能分析

本文对研究区滑坡易发性评价的结果进行了 ROC 曲线下面积(AUC)的验证,以进一步分析模型的精度。AUC 值可以直观地反映模型的合理性,数值越大代表模型结果越合理^[16](图 4、表 4)。WOE 模型的 AUC 值为 0.845 5,而采用 WOE-CatBoost 耦合模型时,AUC 值提升至 0.984 6,总体提升了 0.139 1。这表明在滑坡易发性评价方面,相较于单独采用 WOE 模型,耦合模型能够更精确地预测滑坡的发生风险。

本文还采用混淆矩阵^[17]综合评价模型的性能,采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 值和 Kappa 系数等常用评估指标对模型进行了综合性能评估。结果显示,在准确率方面,耦合模型提升明显,达到 87.28%,这表明耦合模型能够更准确地预测样本的类别。WOE 模型和 WOE-CatBoost 模型的精确率分别为 71.54% 和 93.31%,耦合模型具有稍高的精确率,表明其更能准确地识别正例。在召回率方面,WOE-CatBoost 模型表现更佳,能够更全面地捕捉到正例,而 WOE 模型的召回率相对较低。此外,WOE-CatBoost 模型在 F1 值上表现更出色,说明其在准确率和召回率之间取得了更好的平衡。Kappa 系数反映模型的整体表现,WOE-CatBoost 模型的 Kappa 系数为

0.856 9,而 WOE 模型仅为 0.513 7,这表明耦合模型在整体上优于 WOE 模型。

单一模型和 WOE-CatBoost 耦合模型对于滑坡易发性预测出现较大差异,WOE 是基于概率统计方法,通过计算各个因子的 WOE 值来评估滑坡发生的可能性,忽略了因子之间复杂的交互作用。而 WOE-CatBoost 模型基于 WOE 编码,耦合 CatBoost 利用因子的独立影响,捕捉到这些复杂的因子交互作用,从而提供更准确的预测。

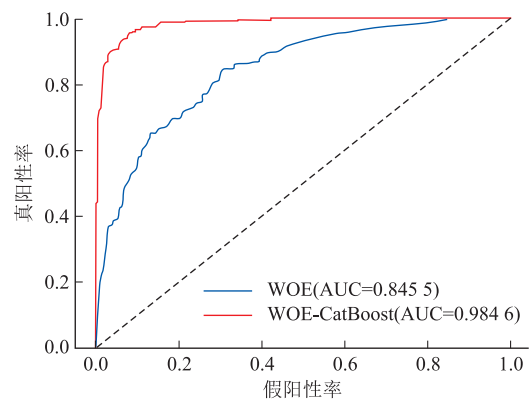


图 4 ROC 曲线图

4.3 结果分析

4.3.1 滑坡易发性结果分析

根据图 3 比较,单一模型与耦合得出的滑坡易

发区存在一定差异,单一模型 5 个易发区域分布不明显,而耦合模型极低易发区域与其他区域分布明

显,但极高和高易发区呈现出相似的规律。极高和高易发区主要集中在研究区的东南部和东北部。在

表 4 WOE 和 WOE-CatBoost 评价指标结果

模型	TP	TN	FP	FN	ACC/%	Recall/%	Precision/%	F1/%	Kappa 系数
WOE	176	498	70	108	79.11	71.54	61.97	66.42	0.513 7
WOE-CatBoost	265	532	36	19	93.54	93.31	88.04	90.60	0.856 9

东北部地区,地形呈现出构造侵蚀剥蚀中山的特征,与马摆河、倘塘河的走向基本一致。而在东南部,地貌主要为间歇性干沟谷,植被稀疏,以宣威-寻甸大断裂为界,其北西属于会泽台褶束,其南东属于曲靖台褶束。易发区主要分布在公路和铁路附近,频繁的人类活动是滑坡灾害发生的主要诱因。地形中的高程和坡度变化使得土壤更容易失稳,增加了滑坡的风险。相反,低易发区主要分布在研究区的中部和西南部,主要为断拗侵蚀溶蚀盆地,走向南北,中部较开阔,南、北较狭窄,阶面平坦,不易发生滑坡灾害。

根据表 5 可知在 WOE 模型中,得到的易发性

区划图极高、高、中、低和极低易发区的面积分别为 1 010.27 km²、1 811.03 km²、1 634.31 km²、1 182.70 km² 和 657.43 km²,滑坡灾害点密度点在极高和高易发区的分别为 0.181 0 个/km² 和 0.074 4 个/km²。当两者耦合时,极高、高、中、低和极低易发区的面积分别为 2 928.96 km²、1 109.37 km²、752.18 km²、677.16 km² 和 828.18 km²,滑坡灾害点密度点在极高和高易发区分别为 0.297 0 个/km² 和 0.023 6 个/km²。由此可见,两者耦合时,虽然极高和高易发区的面积有所减小,但滑坡灾害点密度在极高和高易发区明显提高,显然二者耦合时的分区更加合理。

表 5 滑坡栅格统计结果

易发性等级	WOE 模型						WOE-CB 模型					
	分级面积 /km ²	分级面积 占比/%	滑坡数量 /个	灾害点占 比/%	灾害点密度 /个·km ⁻²	频率比	分级面积 /km ²	分级面积 占比/%	滑坡数量 /个	灾害点占 比/%	灾害点密度 /个·km ⁻²	频率比
极低易发区	1 010.27	16.05	2	0.70	0.002 0	0.04	2 928.96	46.52	2	0.70	0.000 7	0.02
低易发区	1 811.03	28.77	21	7.39	0.011 6	0.26	1 109.37	17.62	5	1.76	0.004 5	0.10
中易发区	1 634.31	25.96	54	19.01	0.033 0	0.73	752.18	11.95	15	5.28	0.019 9	0.44
高易发区	1 182.70	18.79	88	30.99	0.074 4	1.65	677.16	10.76	16	5.63	0.023 6	0.52
极高易发区	657.43	10.44	119	41.90	0.181 0	4.01	828.18	13.15	246	86.62	0.297 0	6.58

4.3.2 评价因子可解释性分析

(1) 全局解释

本文选取 WOE-CatBoost 模型进行 SHAP 解释。从全局解释来看,模型中对滑坡预测影响最大的因子为高程、地层岩性、NDVI、降雨和距水系的距离。图 5 中颜色代表特征值的高低,红色表示高值,紫色表示中间值,蓝色表示低值。SHAP 值大于 0 的点,表示对滑坡产生正向贡献;小于 0 的点则表示对滑坡产生负向贡献。当高程值越小时,对滑坡起正向作用。当 NDVI 处于中间值时对滑坡的发生起正向影响,总体上降雨量为中间值,地层岩性为松散土体岩或者碎屑岩,距水系越近时对滑坡有正向影响。

对滑坡影响因子进行依赖图的绘制,重点关注了对滑坡灾害产生影响的前 5 个因子。对比图 6 和图 7,发现高程与滑坡灾害呈现负相关关系,高程值大于 1 966 m 的范围内,高程对滑坡灾害产生负向影响;而高程值在 1 474~1 966 m 范围内,高程对滑坡灾害产生正向影响,与 WOE 中得到的滑坡发育区段(1 697 m,1 698 m)一致。在地层岩性因子中,碎屑岩对滑坡的影响最为显著,极大增加了滑坡发生的概率,使用 WOE 模型计算的也是碎屑岩对滑坡影响明显。NDVI 对滑坡的影响整体上呈现负相关趋势,NDVI 指数越高,滑坡发生的可能性越小。降雨对滑坡的影响表现为先增加后减少的趋

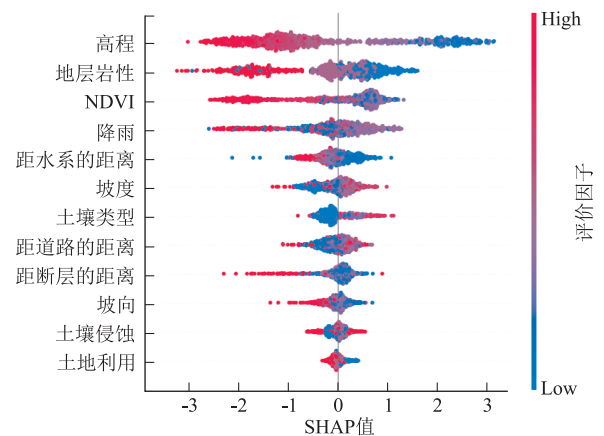


图 5 WOE-CatBoost 模型 SHAP 汇总图

势,当降雨量位于中间值时,滑坡的发生概率增加,趋势与 WOE 模型一致。距水系的距离对滑坡的影响整体呈现负相关趋势,距水系越近,滑坡发生的可能性越大;而当距水系超过 2 000 m 时,滑坡发生的概率显著降低。距水系的距离对滑坡的影响整体呈现负相关趋势,距水系越近,滑坡发生的可能性越大;而当距水系小于 500 m 时,滑坡发生的概率最高。通过 WOE 值计算时,距水系的距离发育区段为(100 m,200 m],可见两者分析结果相似。

(2) 局部解释

选取研究区第 59 个评估单元的滑坡单元,该样本属于极高危险区,预测值为 0.979 6。在这个样本

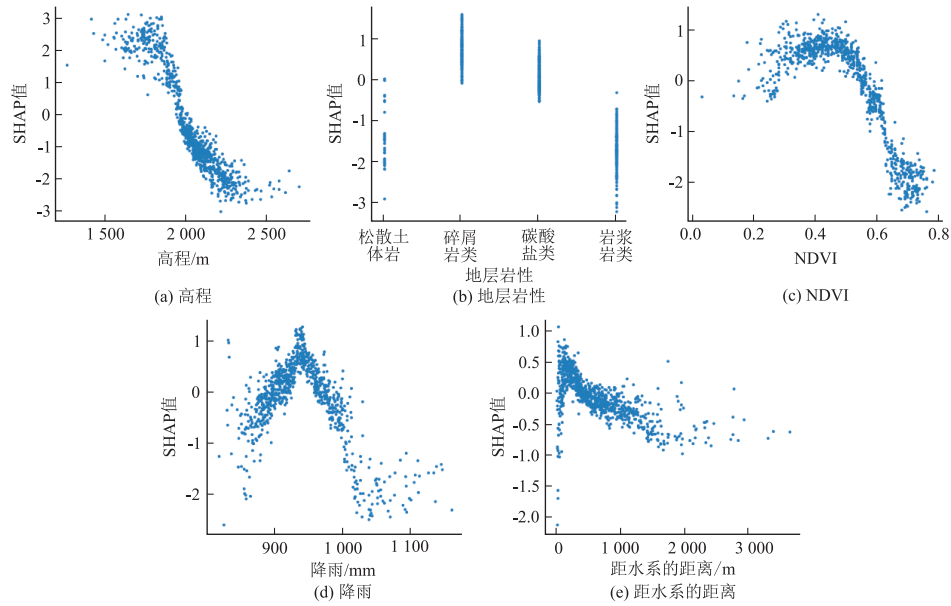


图 6 研究区滑坡灾害重要影响因子依赖图

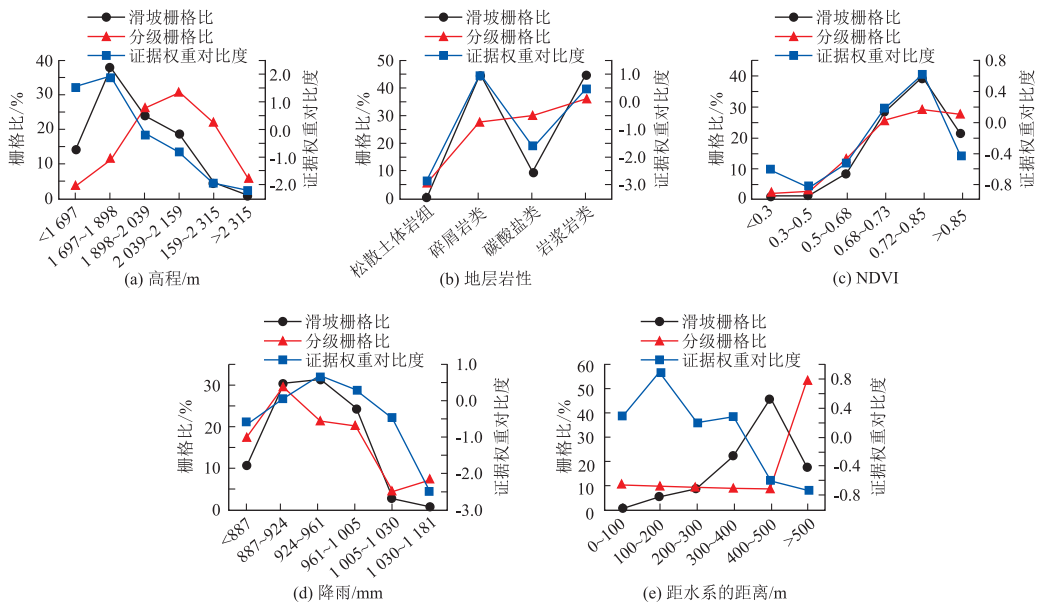


图 7 研究区滑坡灾害重要影响因子分级统计图

中,高程、地层岩性和降雨是对滑坡易发性影响最大的因素,具有较大的正向贡献,对应的 SHAP 值分别为+2.17、+0.84 和+0.76。高程提供滑坡发生的临空面条件,碎屑岩为滑坡提供岩性条件,降雨则为滑坡提供充足的水文条件。相反,坡向和土壤类型是对该滑坡具有负向影响的因素,对应的 SHAP 值分别为-0.2 和-0.12。坡向向西时对滑坡发生产生一定负面影响,而红壤也抑制影响滑坡的发生。

非滑坡样本的选取第 569 个评估单元,其预测值为 0.121 1,属于极低危险区,对于该非滑坡单元,NDVI、高程、坡向、降雨、土壤类型呈抑制作用,该

处植被覆盖率高,降雨量较少,距断层的距离较远抑制了滑坡的发生。通过瀑布图(图 8)的可视化展示,能更清晰地了解各个特征对于该滑坡样本易发性的正向和负向影响,为滑坡发生机理的解释提供了有力支持。

5 结论

本研究通过选择 12 个评价因子,对宣威市的滑坡灾害易发性进行了可解释性研究,并得出以下结论:

- (1) 通过对比 WOE 模型和 WOE-CatBoost 耦

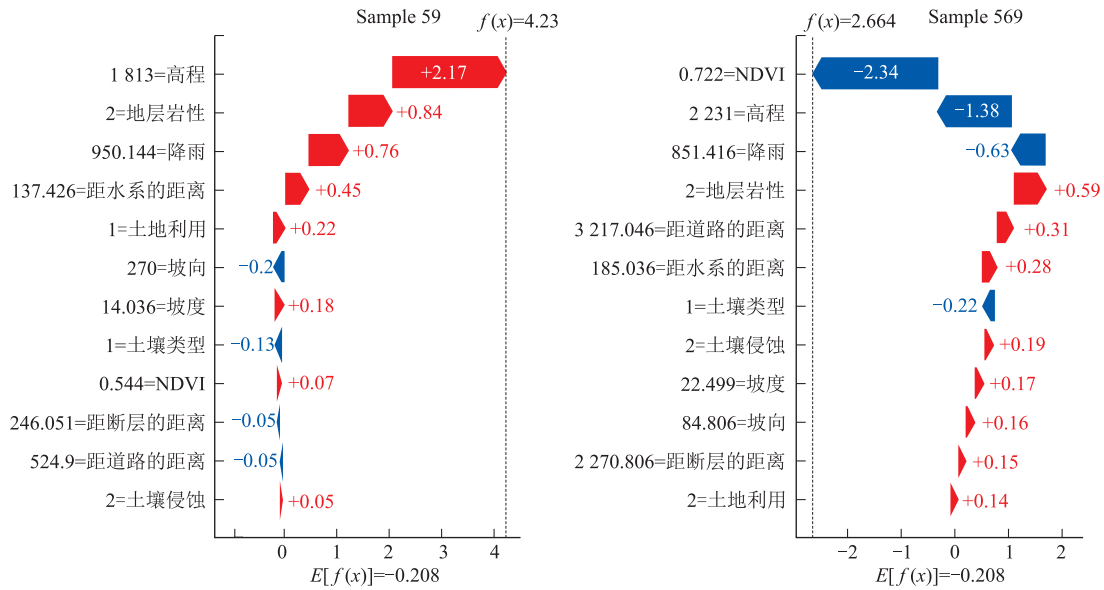


图 8 SHAP 瀑布图

合模型的结果,发现两者对滑坡点的分布和易发区评价结果一致。而在耦合模型中,AUC 值达到了 0.984 6,预测正确率明显优于 WOE 模型,表明其可作为滑坡灾害易发性评价的有效手段。

(2) 采用 WOE 和 SHAP 对研究区进行全局解释,造成该地区滑坡易发的影响因素主要为高程、NDVI、地层岩性和距水系的距离,而降雨是诱发滑坡发生的最主要评价因子。

(3) 滑坡的极高和高危险区域主要分布在研究区的东南部和东北部,低易发区主要分布在研究区的中部和西南部。

参考文献

- [1] 吴宏阳,周超,梁鑫,等.基于 XGBoost 模型的三峡库区燕山乡滑坡易发性评价与区划[J].中国地质灾害与防治学报,2023,34(5):141-152.
- [2] 朱晓曦,王一鸣,龚新法.浙江省文成县滑坡灾害危险性评价[J].中国地质灾害与防治学报,2013,24(3):13-18.
- [3] ADNAN O, TOLGA A. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping; Sultan Mountains, SW Turkey [J]. Journal of Asian Earth Sciences, 2013, 64: 180-197.
- [4] 夏辉,殷坤龙,梁鑫,等.基于 SVM-ANN 模型的滑坡易发性评价——以三峡库区巫山县为例[J].中国地质灾害与防治学报,2018,29(5):13-19.
- [5] 林报嘉,刘晓东,杨川,等. XGBoost 机器学习模型与 GIS 技术结合的公路崩塌灾害易发性研究[J].公路,2020,65(7):20-26.
- [6] HUSSAIN M A, CHEN Z, KALSOOM I, et al. Landslide Susceptibility Mapping Using Machine Learning Algorithm: A Case Study Along Karakoram Highway (KKH), Pakistan[J]. Journal of the Indian Society of Remote Sensing, 2022, 50(5): 849-866.
- [7] ZHANG H, SONG Y, XU S, et al. Combining a class-weighted algorithm and machine learning models in landslide susceptibility mapping: A case study of Wanzhou section of the Three Gorges Reservoir, China [J]. Computers & Geosciences, 2022, 158: 104966.
- [8] 郭子正,殷坤龙,付圣,等.基于 GIS 与 WOE-BP 模型的滑坡易发性评价[J].地球科学,2019,44(12):4299-4312.
- [9] 曾韬睿,王林峰,张俞,等.基于 CatBoost-SHAP 模型的滑坡易发性建模及可解释性[J].中国地质灾害与防治学报,2024,35(1):37-50.
- [10] 邵健,王霄,昌文峰,等.基于 CatBoost 算法的滑坡敏感性模型——以毕节市大方县为例[J].计算机与数字工程,2022,50(6):1365-1370.
- [11] 马祥龙,文海家,张廷斌,等.自动 & 可解释机器学习滑坡易发性评价模型[J].北京师范大学学报(自然科学版):806-818.
- [12] 周新植.滑坡易发性机器学习优化模型及可解释性研究[D].重庆:重庆大学,2022.
- [13] 黄煜,谢婉丽,刘琦琦,等.基于 GIS 与 MaxEnt 模型的滑坡易发性评价——以铜川市中部城区为例[J].西北地质,2023,56(1):266-275.
- [14] 吴常润,赵冬梅,刘澄静,等.基于 GIS 和信息量模型的陇川县滑坡易发性评价[J].西北地质,2020,53(2):308-320.
- [15] 张俊,殷坤龙,王佳佳,等.三峡库区万州区滑坡灾害易发性评价研究[J].岩石力学与工程学报,2016,35(2):284-296.
- [16] 李明,蒋委君,董佳慧,等.基于机器学习的滑坡灾害易发性评价——以三峡库区为例[J].华南地质,2023,39(3):413-427.
- [17] BAHAREH K, BISWAJEET P, AMIR N S, et al. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 49-69.