

DOI:10.13870/j.cnki.stbxb.2026.05.010

CSTR:32310.14.stbxb.2026.05.010

罗登楠,唐杰,谢云,等.东北黑土区农田切沟侵蚀敏感性模型构建与跨流域适用性验证[J].水土保持学报,2026,40(5):104-114.

Luo Dengnan, Tang Jie, Xie Yun, et al. Modeling gully erosion susceptibility on cropland in Northeast China's black soil region: A cross-watershed assessment[J]. Journal of Soil and Water Conservation, 2026, 40(5):104-114.

东北黑土区农田切沟侵蚀敏感性模型 构建与跨流域适用性验证

罗登楠^{1,2}, 唐杰³, 谢云², 谭志成⁴, 成汉铨^{1,2}, 李锴², 徐俊杰^{1,2}

(1.北京师范大学地理科学学部,北京 100875; 2.北京师范大学文理学院,广东 珠海 519087;

3.北京师范大学未来技术学院,广东 珠海 519087; 4.深圳市罗湖外语实验学校,广东 深圳 518004)

摘要: [目的] 针对东北黑土区日益严重的切沟侵蚀问题,利用机器学习模型开展沟蚀风险的精准识别研究及在黑土区不同地理单元下的跨流域泛化能力研究,为制定差异化的水土流失防治策略及保障国家粮食安全提供科学依据与决策支持。[方法] 以鹤北流域和双山流域为研究区,其中鹤北流域用于模型跨流域适用性验证,双山流域为训练区,构建跨流域预测范式。首先,通过文献计量法筛选出文献中使用最多的随机森林(RF)、支持向量机(SVM)和逻辑回归(LR)3种模型;然后,基于地形、水文及人类活动等15个因子构建预测模型,并采用跨流域特大降雨年后验证的方法,系统评估模型性能与跨流域适用性。[结果] 1)随机森林模型表现最优(AUC为0.96),中等敏感等级以上的区域面积占比达78.7%;2)基于2019年与2021年2次特大降雨年的新增切沟验证表明,新增切沟与高风险区吻合区域面积超过52.0%;3)集水区面积、距河流距离与河网密度是核心驱动因子,人类活动与自然因子的交互作用显著促进切沟发育。[结论] 通过引入流域建模和极端事件验证机制,研究结果可为东北黑土区切沟防治与土地可持续管理提供科学依据。

关键词: 东北黑土区; 切沟侵蚀; 敏感性分析; 文献计量法; 机器学习

中图分类号:S157.1;TP79

文献标识码:A

文章编号:1009-2242(2026)05-0104-11

Modeling Gully Erosion Susceptibility on Cropland in Northeast China's Black Soil Region: A Cross-Watershed Assessment

Luo Dengnan^{1,2}, Tang Jie³, Xie Yun², Tan Zhicheng⁴, Cheng Hanquan^{1,2}, Li Kai², Xu Junjie^{1,2}

(1.Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; 2.Faculty of Arts and Sciences,

Beijing Normal University at Zhuhai, Zhuhai, Guangdong 519087, China; 3.School of Technology for Sustainability,

Beijing Normal University, Zhuhai, Guangdong 519087, China; 4.Shenzhen Luohu

Foreign Languages School, Shenzhen, Guangdong 518004, China)

Abstract: [Objective] This study aims to accurately identify gully erosion susceptibility in the black soil region of northeast China using machine learning models. A primary focus is placed on evaluating the cross-watershed transferability of these models, thereby providing a robust scientific basis and decision support for developing conservation strategies and ensuring national food security. [Methods] In response, this paper constructs a cross-watershed prediction framework based on the Hebei and Shuangshan watersheds. The Shuangshan watershed served as the training set, while the Hebei watershed was designated for validating the model's applicability across different catchments. First, three widely used models-Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) were selected based on a bibliometric analysis. Subsequently, gully erosion susceptibility models were established based on 15 factors, including topography, hydrology, and human activities. Their performance and transferability were validated via post-hoc validation under cross-watershed

收稿日期:2026-02-05

修回日期:2026-03-17

录用日期:2026-03-23

网络首发日期(www.cnki.net):2026-07-06

资助项目:国家重点研发计划项目(2024YFD1501200);国家自然科学基金青年基金项目(42307425);国家自然科学基金面上项目(42477342)

第一作者:罗登楠(1997—),男,博士研究生,主要从事土地退化研究。E-mail:luodn@mail.bnu.edu.cn

通信作者:唐杰(1992—),男,博士,副研究员,主要从事土壤侵蚀机理研究。E-mail:jiatang@bnu.edu.cn

http://stbxb.alljournal.com.cn

extreme rainfall years. [Results] 1) The RF model performs optimally ($AUC=0.96$), with the area proportion of regions above the moderate sensitivity level reaching a total of 78.7%; 2) Verification based on newly formed gullies from two extreme rainfall years (2019 and 2021) demonstrates that the overlap between newly formed gullies and high-sensitivity areas exceeds 52.0%; 3) Catchment area, distance to rivers, and river network density are key driving factors. Furthermore, the interaction between human activities and natural factors significantly promotes gully development. [Conclusion] This research establishes a scientific framework for gully erosion control and sustainable land management in the black soil region of northeast China through cross-watershed modeling and validation under extreme rainfall years.

Keywords: northeast black soil region; gully erosion; sensitivity analysis; bibliometric analysis; machine learning

Received: 2026-02-05

Revised: 2026-03-17

Accepted: 2026-03-23

Online(www.cnki.net): 2026-07-06

东北黑土区是中国主要的粮食基地,虽然开垦时间短,但因受到季节性的冻融作用、春季融雪和夏季集中性降雨、长缓坡地形,以及人类不合理的耕种活动等综合影响,区域内农地土壤侵蚀十分严重。切沟侵蚀不仅导致耕地面积减小、破碎化程度加剧、耕作成本增加,还加剧下游河道、湖泊和水库等的泥沙淤积,增大洪水发生风险。水利部2023年公布的东北黑土区侵蚀沟调查数据^[1]显示,东北黑土区的侵蚀沟总数超过66.67万条,且大多处于活跃期,开展切沟侵蚀的敏感性评估,识别高风险区域对稳定农业生产、保障粮食安全具有重要的意义。

切沟侵蚀机理复杂,受到气候、土地利用、地形和人类活动等多种因素的综合影响^[2]。地形是控制切沟侵蚀的关键因素,切沟长度、侵蚀体积与坡面长度和汇水面积等地形因子呈显著相关而与坡度呈负相关。人类活动引起的土地利用变化对切沟的发育产生了重要影响^[3]。起垄种植中不同垄作区切沟侵蚀强度差异明显。由于道路路面入渗率低,降雨会迅速汇集并沿道路流动,从而增加下游沟蚀风险^[4]。防护林通过降低风速、截留降水、增加土壤下渗及优化土壤结构等方式减少径流量和降低径流侵蚀能力,对抑制切沟侵蚀具有积极作用^[5-6]。但顺坡的林带可能导致水流汇集,增加沟蚀^[7]。人类活动虽已被证实对切沟发育有显著影响,但在切沟敏感性研究中对人类活动因子的考量尚不充分,更缺乏自然因素和人为因子的交互作用分析。因此,将这些关键人为因子纳入切沟敏感性评估体系,并解析其与自然因子的耦合效应是提升切沟敏感性模型的区域适用性和预测精度的关键。

传统方法主要是运用统计学模型。例如,Arabameri等^[8]在伊朗对比频率比(frequency ratio, FR)、证据权重(weight of evidence, WoE)、熵指数(index of entropy, IoE)和层次分析法(analytic

hierarchy process, AHP) 4种传统模型发现,层次分析法的效果是最准确的。机器学习的出现为切沟侵蚀风险评估提供了新的可能。Pourghasemi等^[9]根据人工神经网络(artificial neural network, ANN)、支持向量机(support vector machine, SVM)和最大熵模型(maximum entropy model, MaxEnt)及两两组合的集成模型运用到伊朗阿格马姆流域切沟敏感性评估表明,人工神经网络和支持向量机组成的集成模型预测效果最好;Huang等^[10]在东北黑土区克山县引入10个与切沟侵蚀相关的地形因子,使用随机森林模型制作当地的切沟侵蚀敏感区地图,取得良好的效果。然而,针对不同区域切沟敏感性评估研究,最优模型也有所不同,频率比^[8]、熵指数^[8]、卷积神经网络^[11]等模型均在相应的研究中呈现较好的切沟敏感性预测结果。针对东北黑土区,尚缺乏借助文献调研,系统评估和对比不同模型在东北黑土区切沟敏感性预测中的适宜性,确定最优的预测模型。此外,大多数研究的模型验证多局限于同一流域内的静态样本;模型的泛化能力评估缺乏,无法检验跨地理空间单元的外推能力,即用一个流域训练好的模型去预测另一个独立流域的切沟侵蚀风险,使得模型跨区域应用与实践指导意义大打折扣。模型能否在特大降雨年下保持其预测的可靠性,是其能否真正服务于切沟侵蚀防治的核心。而现有研究鲜有利用特大降雨年份新增的切沟作为独立的后验验证集,来检验切沟敏感性预测的实际效果和模型的泛化能力,难以验证模型在真实降雨事件下的预测能力。

因此,本研究以东北黑土区的鹤北流域和双山流域为研究区,旨在解决的关键问题为:1)结合文献计量法,探究何种机器学习模型在该区域的切沟侵蚀敏感性分析中适宜性最好;2)综合考虑地形、水文及道路、防护林和耕作方式等关键人类活动相关因子,并探究自然与人为因子的交互作用对切沟侵蚀

的贡献;3)创新性地采用“跨流域外推”验证范式,检验最优模型在不同流域间的空间泛化能力;4)利用2019年与2021年2次特大降雨年份新增的切沟实测数据,对模型的空间外推结果进行独立的后验验证,评估其在极端事件下的预测稳健性。本研究不仅为东北黑土区提供一种新的、可靠的切沟敏感性评估思路,也期望通过对模型、因子及验证方法的系统性探究,深化对黑土区切沟侵蚀发生机理的科学认知,为东北黑土区水土保持的精准施策与可持续发展提供有力的科技支撑。

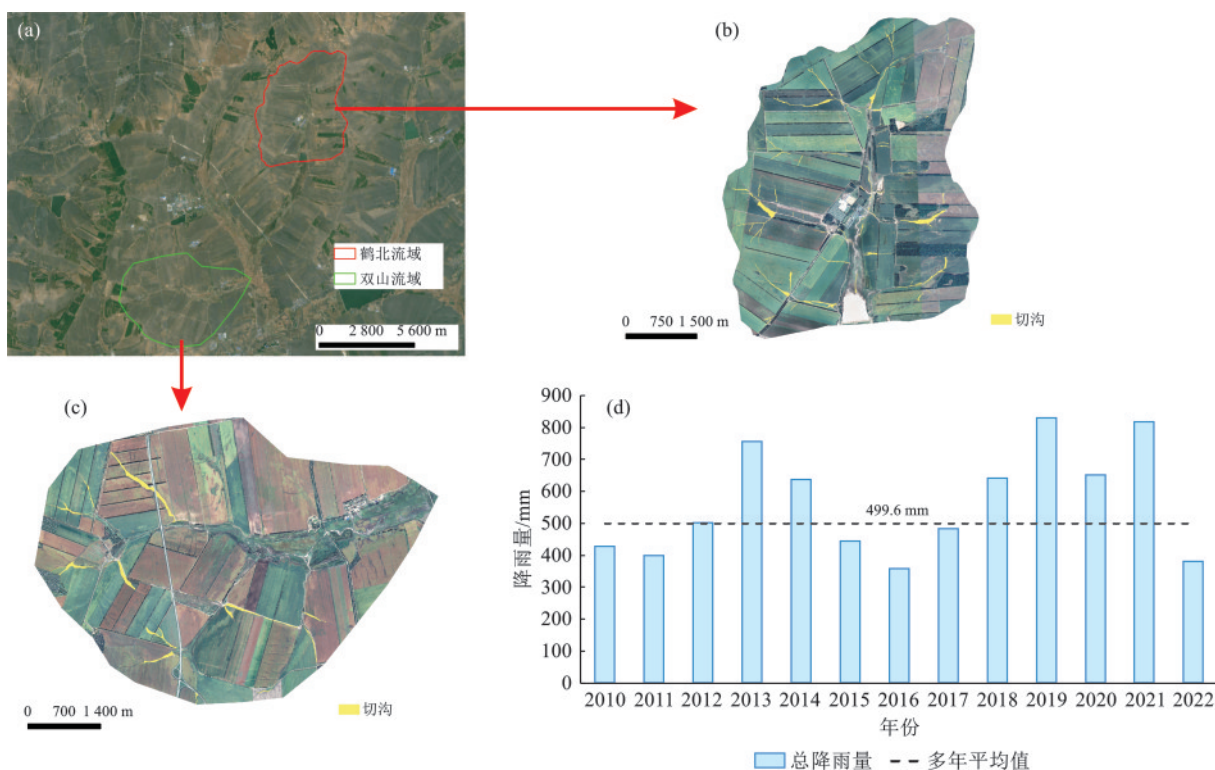
1 材料与方法

1.1 研究区概况

选取的研究区为鹤北流域($48^{\circ}59' \sim 49^{\circ}03'N$, $125^{\circ}16' \sim 125^{\circ}21'E$)和双山流域($48^{\circ}53' \sim 48^{\circ}55'N$, $125^{\circ}09' \sim 125^{\circ}16'E$),面积分别为28.02、29.30 km²。2个流域均隶属于黑龙江省嫩江市鹤山农场,由北大

荒集团九三分公司管理,都是以种植大豆、玉米为主的典型现代化机耕小流域。

2个流域实际相距13 km,均隶属于嫩江流域水系。它们位于东北典型的漫川漫岗丘陵区,坡缓坡长是其典型的地形特征,坡度为 $1^{\circ} \sim 6^{\circ}$,坡长为500~4 000 m。研究区属于温带季风气候,冬季寒冷漫长,无霜期短,年平均气温 $-7 \sim 11^{\circ}C$;夏季高温多雨,平均气温 $19.6^{\circ}C$ 。鹤山农场气象站资料显示,1976—2022年多年平均降水量为499.6 mm,其中特大降雨年2013年、2019年和2021年的降雨量分别为754.6、829.0、816.4 mm(图1d)。研究区土壤以黑土为主,质地黏重,透水性差,很容易在降雨径流的作用下发生水力侵蚀。斜坡和顺坡起垄耕作普遍,近年来侵蚀沟发育迅速,严重影响农业发展^[12]。流域的相对位置见图1。2021年解译结果显示,鹤北流域和双山流域切沟分别为142条和183条。



注:(a)为鹤北流域(红色区域)和双山流域(绿色区域)的相对位置;(b)为鹤北流域切沟分布;(c)为双山流域切沟分布;(d)为2010—2022年嫩江年降雨量,虚线表示1976—2022年多年平均降雨量,为499.6 mm。

图1 鹤北流域与双山流域位置概况

Fig.1 Geographical locations and overview of the Hebei and Shuangshan watersheds

1.2 基础数据准备

1.2.1 数据收集与处理 本研究使用的基础数据主要是谷歌影像数据(<https://map.wemapgis.com/>)和ALOSDEM数据(Alaska Satellite Facility, <https://search.asf.alaska.edu/#/dataset=ALOS>)(表1)。相比传统的30 m分辨率DEM,ALOSDEM数据经过重采样与地形精校正,能够提供更平滑的坡度梯度

信息,有效减少由于卫星轨道抖动产生的系统性条纹噪声。

根据多年降雨资料显示,2013年、2019年和2021年均均为特大降雨年(图1d),因此,选择2013年、2019年和2021年的遥感影像进行切沟解译,其中鹤北流域提取2013年、2019年和2021年的切沟矢量数据,双山流域仅解译提取2021年的切沟矢量数据。

表 1 基础数据

Table 1 Summary of datasets used in this study

数据类型	数据名称	分辨率/m	年份	主要用途
遥感影像	Google Earth 影像	0.595	2013	侵蚀沟的解译提取
			2019	
			2021	
地形数据	ALOS-DEM	12.500	—	因子数据提取、发生风险评估

双山流域 2021 年切沟矢量数据主要用于构建切沟机器学习模型的训练和验证数据集。双山区域的切沟面积相对于流域面积相差悬殊,若使用完全随

机点选取法所取得的数据集样本可导致属于切沟的样点极少,因此,在切沟范围内和非切沟范围内各随机选取 1 500 个样点,共 3 000 个样点,构成双山流域训练数据集。数据集中的 80% 用于机器学习模型的训练,20% 用于机器学习模型的验证。

1.2.2 切沟样本数据统计分析 本研究对 2 个流域进行了切沟的解译,得到双山流域 2021 年切沟矢量数据和鹤北流域 2013 年、2019 年和 2021 年切沟矢量数据。同时,还获取鹤北流域在特大降雨年 2019 年相对于 2013 年新增的切沟矢量数据和 2021 年相对于 2019 年新增的切沟矢量数据。具体切沟特征见表 2。

表 2 切沟特征统计分析

Table 2 Statistical analysis of gully characteristics

流域名称	年份	数量/条	长度/m	平均长度/m	面积/m ²	平均面积/m ²
双山流域	2021	183	19.5~1 268.0	236.4	87.6~31 838.6	4 238.9
	2013	192	16.7~1 020.0	182.4	24.5~52 051.2	1 779.2
	2019	79	36.5~1 020.0	242.0	1 580.0~52 051.2	3 068.8
鹤北流域	2021	142	14.9~1 061.7	200.6	29.1~52 051.2	2 611.5
	2019	113	16.7~733.6	148.9	24.5~17 376.8	882.3
	2021	63	14.9~656.7	176.2	29.1~28 773.9	1 761.1

1.2.3 切沟侵蚀敏感性预测指标体系构建 根据现有涉及影响切沟侵蚀发生的相关因子研究及多重共线性检验的结果^[13],选择地形、水文和人类活动 3 大类指标作为切沟侵蚀发生风险评估的因子。地形因子包括海拔、坡向、坡度、平面曲率、剖面曲率、坡度坡长因子和地形起伏度;水文因子包括河网密度、地形湿度指数、水流动力指数和距河流距离;人类活动因子包括距道路距离、距林带距离和耕作方式。

海拔是影响切沟侵蚀的重要因素之一,影响产流过程^[14];坡向通过影响降雨、风、阳光和流域形态结构影响切沟侵蚀^[15],其中坡向正北为 0°,顺时针变化角度变大,依此类推出坡向的分类,分别为平地、北、东北、东、东南、南、西南、西、西北 9 类;坡度是影响切沟侵蚀的关键因素,影响径流、排水强度及土壤颗粒的分离^[10];平面曲率和剖面曲率是切沟侵蚀的重要预测因子,对切沟侵蚀过程的影响在于控制水流下坡时的发散或收敛^[16]。理论上,山坡在平面上的曲率被定义为:正值表示下坡方向的凸度,负值表示下坡方向的凹度,如果值为 0,则为平面^[17]。在剖面曲率输出中,如果值为负,则说明该像元的表面向上凸起;如果剖面曲率为正,则说明该像元的表面开口朝上凹入^[18]。

坡度坡长因子(LS)指地表径流的速度,被认为是最重要的泥沙输移因子之一。计算公式^[19]为:

$$LS = \left(\frac{f \times \text{cellsize}}{22.13} \right)^{0.4} \times \left(\frac{\sin \beta}{0.0896} \right)^{1.3} \quad (1)$$

式中: f 为汇流累积量,无量纲; β 为坡度,(°);cellsize 为栅格数量。

地形起伏度指数(TR)用于捕捉冬、夏降雨期间地形局部效应对切沟侵蚀发生间接影响的更适宜的自变量,计算公式^[20]为:

$$TR = \sum (Z_{ij} - Z_{00})^2 \quad (2)$$

式中: Z_{ij} 为中心像元各相邻像元的高程,m; Z_{00} 为中心像元的高程,m。

河网密度(DD)指单位面积内的河流长度,较高的河网密度会增加地表径流,从而增加切沟侵蚀发生的可能性。计算公式为:

$$DD = \frac{L}{A} \quad (3)$$

式中: L 为区域内河流的总长度,m; A 为该区域的面积,m²。

集水区面积是沟蚀发育的重要影响因子。该因子可用于确定地表降雨径流、地表径流和地下径流的空间分布特征,并识别潜在的侵蚀发生点^[21]。

地形湿度指数(TWI)可用于识别降雨径流模式、土壤水分体积分数潜在增加区域及积水区,并能预测集水区临时性沟壑的位置^[18]。

$$TWI = \ln \left(\frac{A_s}{\tan \alpha} \right) \quad (4)$$

式中: A_s 为特定集水区面积,m²/m; α 为坡度,(°)。

水流动力指数(SPI)能有效反映沟蚀易发区域的空间特征。计算公式^[19]为:

$$SPI = A_s \tan \alpha \quad (5)$$

距河流距离与河网密度共同调控坡面径流量,其空间异质性直接影响沟蚀发生的临界水动力条件。有研究^[22]表明,二者对沟蚀发生的敏感性具有显著正相关。

距道路的距离是沟蚀风险分析的控制性因子。道路工程会加剧地表径流汇集,并将其分散至其他小流域,从而促进侵蚀沟道的发育^[22]。相关研究^[23]表明,受林地与防护林带影响的耕地坡面,其季节性沟蚀的体积土壤流失量较无林地影响的坡面高出2倍以上,距林带的距离是一个不可忽视的要素。耕作方式通过改变地表粗糙度和影响径流模式,对沟蚀的发生和发展具有重要影响。耕作方式通过确定耕地垄向与坡向的夹角来实现。在1块完整的耕地中,如果夹角为 $0^\circ \sim 20^\circ$,则为顺坡耕作;如果夹角为 $20^\circ \sim 70^\circ$,则为斜坡耕作;如果夹角为 $70^\circ \sim 90^\circ$,则为横坡耕作^[24]。

本研究使用 Google Earth 影像和 ALOSDEM 数据,在 ArcGIS 10.8 软件中计算得到双山区域和鹤北流域各相关要素指标的空间分布图,并通过重采样的方式获得 $10\text{ m} \times 10\text{ m}$ 栅格单元各相关因子指标的空间分布图。

1.3 研究方法

1.3.1 文献计量法 通过 Web of Science 收集相关文献,并对收集到的文献进行预处理,包括数据清洗、关键词提取等步骤;利用 Citespace 软件构建关键词共现网络图及关键词突现图,统计出使用机器学习对切沟侵蚀敏感性进行研究的相关文献中使用频率最高的3种机器学习模型。

1.3.2 模型选择与性能评估 根据文献调研采用逻辑回归(logistic regression, LR)、随机森林(random forest, RF)和支持向量机(support vector machine, SVM)3种不同的机器学习模型进行切沟敏感性的预测。逻辑回归在切沟侵蚀敏感性评价中是一种广泛应用的非线性多元统计模型^[25]。它利用最大似然估计法,能够保证每一点的拟合为最优。随机森林通过聚合多个决策树的预测,有效缓解多重共线性问题,提高模型的准确性和可解释性^[26]。支持向量机能够在样本较少的情况下较好地解决非线性及高维模式识别问题,已广泛用于趋势分析和预测分析中。

为评估模型性能稳健性,通过绘制 ROC 曲线(receiver operator characteristic curve, ROC),计算 AUC (area under the roc curve, AUC)对切沟侵蚀敏感性的预测模型进行评估。在二分类模型性能评估中,ROC 曲线分析作为核心方法学工具,通过构建不同阈值下模型性能的表现来实现模型验证。其数学

原理体现为:以模型对侵蚀发生单元的正确识别率为纵坐标,对非侵蚀单元的错误判定率为横坐标,在 $[0, 1]$ 概率空间内生成连续决策边界曲线。通过积分运算获得的曲线下面积(AUC),其理论值域界定为 0.5(随机判别基线)至 1.0(理想判别能力)^[26],当 $AUC \in (0.7, 0.8]$ 时表示模型具有实用价值, $AUC > 0.8$ 则达到空间预测优等级别。ROC 曲线几何形态与左上顶点的趋近程度直观表征模型判别精度。

1.3.3 单因子和交互因子重要性排序 为评估单因子预测变量对切沟敏感性预测的相对重要性,通过由实际值和预测值之间的均方误差增加的百分比(%IncMSE)来量化。%IncMSE 值越大,表示变量越重要。

为量化环境因子相互作用对切沟侵蚀的影响程度,通过构建交互增强型随机森林模型解析相关因子对切沟侵蚀敏感性的影响机制。首先,对 15 个因子进行二阶交互项构造,生成包含两两因子乘积项的 105 个交互因子,加上单因子共 120 个加入模型的因子,以捕捉非线性协同效应;其次,采用弹性网络正则化策略优化模型参数,通过交叉验证确定最佳正则化强度与类型,抑制多重共线性干扰;最后,基于最优模型的标准化回归系数绝对值量化特征贡献度,其中主效应表征单一因子的独立作用,交互效应揭示因子间耦合关系,通过贡献度权重排序提取单因子(如河网密度、距河流距离)及其交互因子(如耕作方式 $\times$ 坡度),实现各交互因子的贡献强度的量化,为解析切沟发育的多尺度驱动机制提供量化依据。

1.4 技术路线

基于双山流域的遥感影像数据、数字高程模型(DEM)数据及切沟矢量数据,构建该区域的样本点数据集,用于机器学习模型的训练、验证和参数优化;通过模型评估获取各算法的 ROC 曲线、AUC 值及 15 个影响因子的单因子与交互因子重要性排序结果;最后,将训练完成的切沟侵蚀敏感性预测模型应用于鹤北流域,生成该区域的切沟侵蚀敏感性空间分布图(图 2)。

2 结果与分析

2.1 基于文献计量法的机器学习模型选取

利用 Web of Science 的高级检索,在 SCI 核心数据库中 进行 检索 发现,2000—2025 年,在“machine learning”+“gully erosion”的关键词下搜索到 168 篇文献。利用 Citespace 软件进行可视化分析,得出关键词共现图(图 3)和关键词频次排序表(表 3)。在关键词共现图中,圆形越大,表明该关键词出现的频率越大。在 168 篇文献中,属于机器学习模型的关键词排名前 3 的为“logistic regression(逻辑回归)”“Random Forest(随机森林)”和“Support vector machine(支持向量机)”。

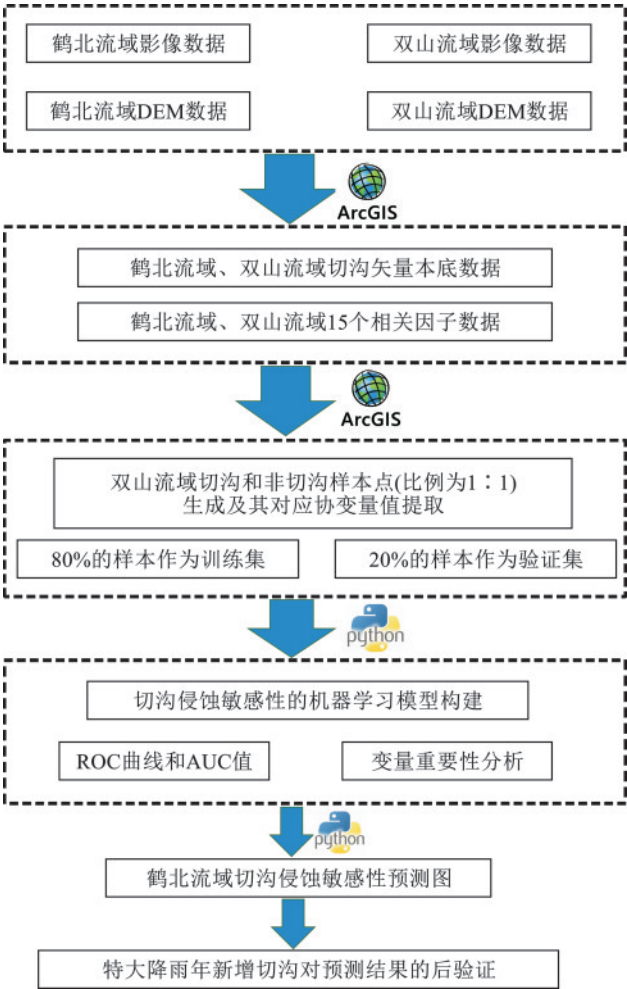


图 2 技术路线
Fig.2 Research framework

在关键词频次排序表中,纵列第 1 行为关键词出现的频次(降序排列),纵列第 2 行为关键词。在 168 篇文献中,属于机器学习模型的关键词排名前 3 的为出现 75 次的“logistic regression(逻辑回归)”、出现 51 次的“Random Forest(随机森林)”和出现 26 次的“Support vector machine(支持向量机)”。

在 CNKI 数据库的检索结果显示,在“机器学习”和“侵蚀”这 2 个关键词下,2000—2024 年仅有 24 篇文献,参考价值较低,因此不纳入分析范畴。综上,本研究利用文献综述法选取得到目前所有使用机器学习对切沟侵蚀进行研究的文献中,使用频率最高的 3 种机器学习模型为逻辑回归(LR)、随机森林(RF)和支持向量机(SVM)。因此,本研究将使用这 3 种机器学习模型进行切沟侵蚀敏感性的研究。

2.2 模型性能评估与因子重要性排序

3 种模型在训练流域中得到的 ROC 曲线特征及 AUC 值见图 4。结果表明,3 种模型的 AUC 值均显著高于 0.80 的良好阈值,其中逻辑回归模型的 AUC 值为 0.85,支持向量机模型的 AUC 值为 0.89,而随机森林模型的 AUC 值高达 0.96。可知本研究构建的 3 种机器学习模型均能有效实现切沟侵蚀敏感性的精准预测,满足区域切沟侵蚀风险评估的技术需求。从模型性能对比来看,随机森林模型表现最优,其 AUC 值(0.96)相较于逻辑回归与支持向量机分别提升 12.9% 和 7.9% 的模型性能,远高于 0.90 的优异阈值,表明该模型在切沟侵蚀敏感性分类中具有极高的预测精度与可信度。



图 3 SCI 数据库下的关键词共现图
Fig.3 Co-occurrence network of keywords indexed in the SCI database
<http://stbcbx.alljournal.com.cn>

表 3 SCI 数据库下的关键词频次排序
Table 3 Ranking of keyword frequencies based on the SCI database

出现频次	关键词(英文)	关键词(中文)	重要度	中心度
75	soil erosion	土壤侵蚀	16	0.02
75	logistic regression	逻辑回归	9	0.02
74	gully erosion	切沟侵蚀	15	0.03
57	machine learning	机器学习	13	0.04
51	random forest	随机森林	10	0.04
42	landslide susceptibility	滑坡敏感性	14	0.04
29	GIS	地理信息系统	13	0.05
29	prediction	预测	11	0.05
26	support vector machine	支持向量机	14	0.06
25	machine learning models	机器学习模型	16	0.07
25	gully erosion susceptibility	切沟侵蚀敏感性	15	0.04
22	classification	分类	11	0.05
22	region	区域	10	0.01
21	land use	土地利用	22	0.11
21	performance	表现	12	0.04
21	models	模型	4	0
20	semiarid region	半干旱地区	23	0.12
20	spatial prediction	空间预测	13	0.07
18	frequency ratio	频率比例	19	0.11
17	land degradation	土地退化	18	0.10
16	regression	回归	6	0.01
15	certainty factor	确定性因素	17	0.06
15	susceptibility	脆弱性	16	0.09
15	data mining techniques	数据探勘技术	11	0.06
15	river basin	流域	11	0.02

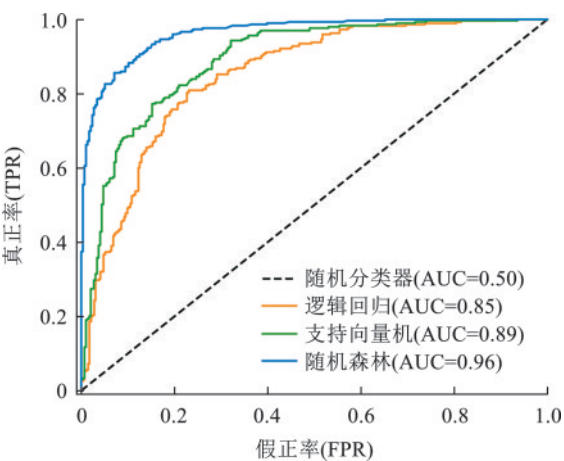
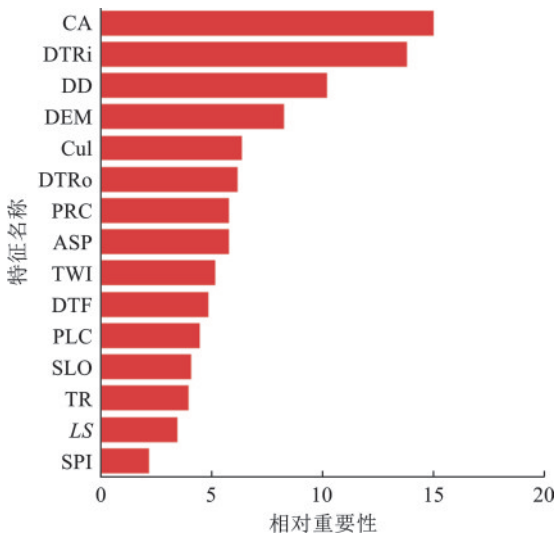


图 4 3 种切沟侵蚀敏感性的机器学习模型的 ROC 曲线
Fig.4 ROC curves of the three machine learning models used for assessing gully erosion susceptibility

在逻辑回归(LR)、随机森林(RF)和支持向量机(SVM)模型中,随机森林(RF)模型的 ROC 曲线和 AUC 值表现最优,因此,在变量重要性分析中仅考虑随机森林模型。针对随机森林(RF)模型所纳入的 15 个影响因子的重要性排序结果表明,各因子均对切沟侵蚀发生具有显著影响。其中,集水区(CA)、距河流距离(DTRi)及河网密度(DD)的重要性位居前 3 位;而耕作因子[距道路距离(DTRo)、耕作方式(Cul)、距离林带距离(DTF)]在模型中也体现出对切沟侵蚀的显著作用,为揭示人类活动对切沟侵蚀的驱动机制提供关键支撑(图 5)。



注:CA为集水区;DTRi为距河流距离;DD为河网密度;DEM为海拔;Cul为耕作方式(垄向与坡向的夹角);DTRo为距道路距离;PRC为剖面曲率;ASP为坡向;TWI为地形湿度指数;DTF为距离林带的距离;PLC为平面曲率;SLO为坡度;TR为地形起伏度;LS为坡长坡度因子;SPI为水流动力指数。下同。

图 5 随机森林模型单因子重要性排名
Fig.5 Relative importance of individual factors based on the Random Forest (RF) model

交互因子为相关因子两两组合形成的新变量,用于提升切沟侵蚀敏感性预测的准确性。本研究共选取 15 个相关因子,两两组合可生成 105 个交互因子;因数量较多,仅呈现随机森林模型中重要性排名前 15 的交互因子(图 6)。结果显示,随机森林模型中交互因子重要性居前 3 位的依次为河网密度×集水区、平面曲率×距河流距离和坡度×距河流距离。值得注意的是,交互因子的重要性排名与单因子高度一致,即与距河流距离、集水区及河网密度相关的交互因子仍显著靠前。部分涉及人类活动(如距道路距离、距林带距离和耕作方式)的交互因子,其重要性相较于对应的单因子得到显著提升。这些重要性获得跃升的交互因子均为“人类活动-环境因子”的耦合类型。具体来说,与距林带距离、耕作方式和距道路距离组合的交互因子,其重要性排名分别位列

第 9、11、12 位。研究表明尽管单一的人类活动因子对切沟侵蚀的直接影响可能不显著,但当其与地形、水文等关键环境要素发生耦合作用时,可对切沟侵蚀过程产生重要的、不可忽视的调控效应。

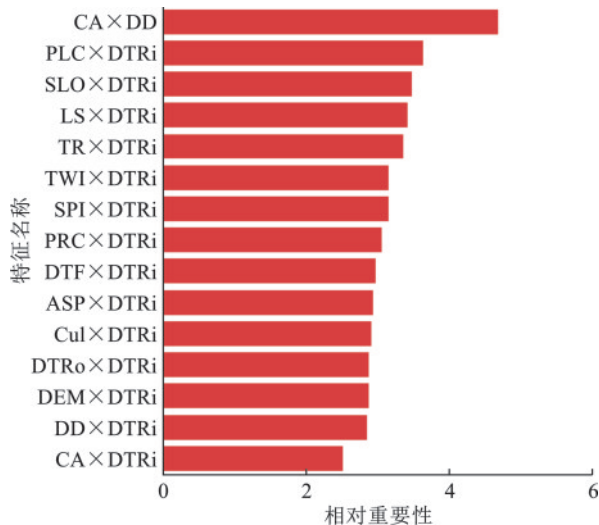


图 6 随机森林模型交互因子重要性排名
Fig.6 Relative importance of interactive factors based on the Random Forest (RF) model

2.3 切沟侵蚀敏感性空间分布特征

2.3.1 切沟侵蚀敏感性预测结果 基于前文验证合格的逻辑回归(LR)、随机森林(RF)与支持向量机(SVM)模型,生成鹤北流域 10 m 分辨率的切沟侵蚀敏感性预测图(图 7)。切沟侵蚀敏感等级划分采用 ArcGIS 10.8

的自然断点法,该方法通过最大化类间差异和最小化类内差异,能客观反映敏感性值的自然分布特征,将切沟侵蚀敏感性划分为低、中、高和极高 4 类^[24]。

对比各模型的敏感性预测(图 7a~图 7c)发现,3 个模型预测的高敏感区具有极高重合度,相互验证说明可信度与真实度较高。各敏感等级区域面积占比图表明,在随机森林模型中,低敏感区、中敏感区、高敏感区和极高敏感区分别占 21.3%、31.6%、31.0% 和 16.1%,中等敏感等级以上的区域面积占比合计达 78.7%,说明鹤北流域的大多数区域目前面临的切沟侵蚀风险不可忽视。

结合鹤北流域的切沟分布与敏感性关系(图 7e)发现,在逻辑回归模型中,低敏感区和极高敏感区的切沟率分别为 7.3% 和 44.7%;在随机森林模型中,低敏感区和极高敏感区的切沟率分别为 9.6% 和 21.0%;在支持向量机模型中,低敏感区和极高敏感区的切沟率分别为 8.4% 和 39.5%,即切沟分布区域与切沟侵蚀敏感性高值区具有高度的重合,进一步验证机器学习模型的“跨流域”预测的准确性和可靠性;同时存在部分切沟侵蚀敏感性较高但尚未发生切沟侵蚀的区域,但并不说明模型对于切沟侵蚀的预测不准确。事实上,说明在切沟侵蚀的高敏感区域,虽然切沟侵蚀尚未形成,但未来极有可能在该区域形成新的切沟。因此,识别此类区域,对于指导切沟侵蚀的防治更为重要。

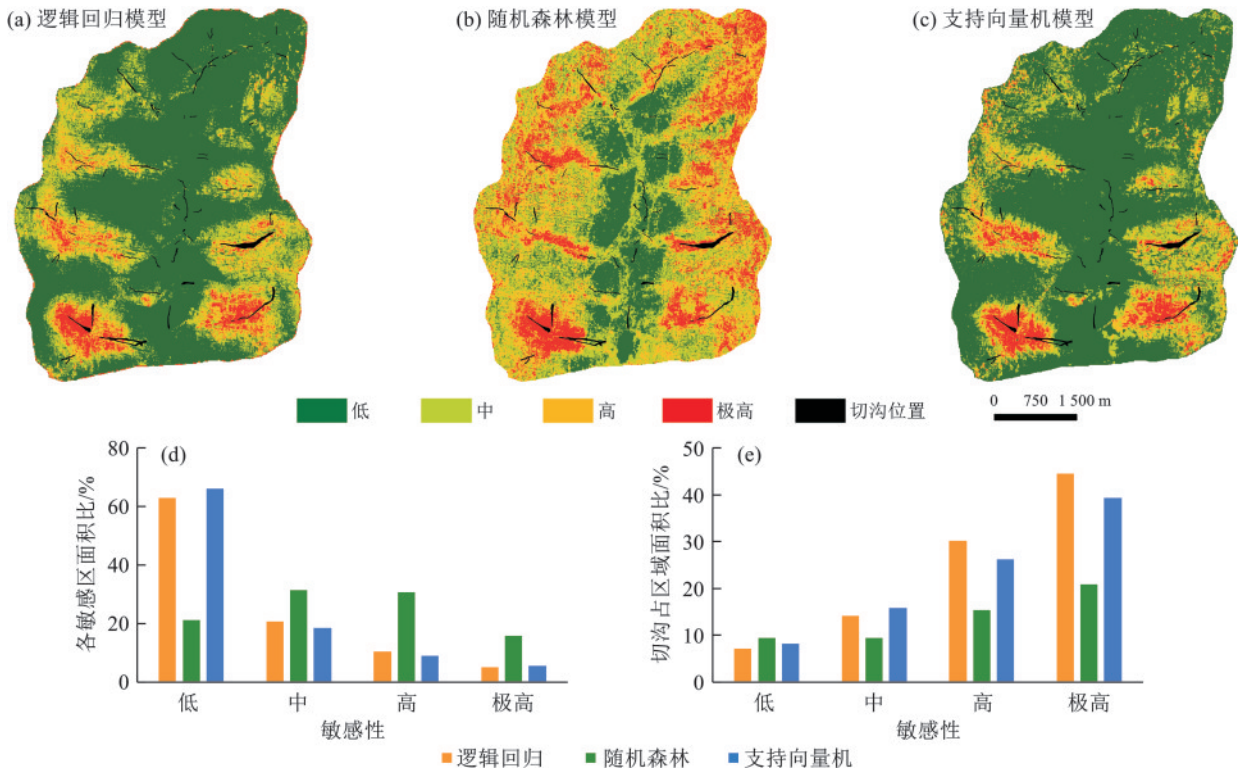


图 7 切沟侵蚀敏感性空间分布特征
Fig.7 Spatial prediction of gully erosion susceptibility and statistical analysis across different models
<http://stbcbx.alljournal.com.cn>

2.3.2 鹤北流域切沟变化与敏感性关系 为对切沟敏感性评价模型的预测精度及时空应用的稳健性进行严格检验,本研究引入2个独立时期的特大降雨年(2019年和2021年)所诱发的新增切沟数据,即2019年相较于2013年新增切沟和2021年相较于2019年新增切沟作为对模型的后验验证集。

在2019年特大降雨年的检验中,随机森林模型展现出优异的预测准确性(图8)。结果显示,高达52.57%(高敏感区30.13%+极高敏感区22.44%)的新增切沟精准地分布于模型划定的高风险区域内,而仅有17.29%位于低敏感区,初步证实RF模型能够有效识别切沟侵蚀的关键控制因子,其预测结果与特大降雨年份下的切沟侵蚀现实情况高度吻合。相比之下,逻辑回归(LR)和支持向量机(SVM)模型则表现出显著偏差,分别将48.69%和55.75%的实际侵蚀点错误归为低敏感区,在防灾实践中可能导致对沟蚀风险的严重低估。

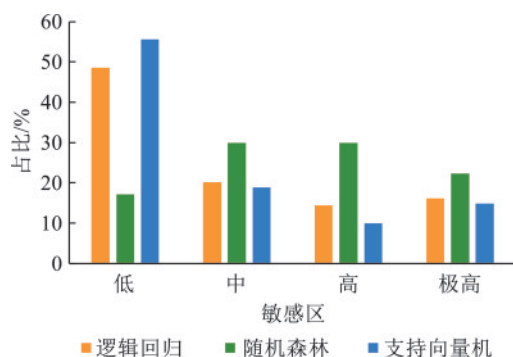


图8 2019年相对于2013年的新增切沟在各等级敏感区域内的占比

Fig.8 Proportional distribution of new gullies (2019 vs. 2013) across different gully erosion susceptibility classes

在2021年的特大降雨年中,新增切沟在随机森林模型预测的高、极高敏感区内新增切沟的占比达60.94%,其中极高敏感区和高敏感区分别占38.45%和22.49%(图9)。而位于低敏感区的比例在3个模型中最低,仅为15.98%。表明随机森林模型并非仅对特定年份的特大降雨年有效,而是具有高度的时间稳健性和预测一致性,充分体现模型“跨流域”的时间泛化能力。随机森林模型成功捕捉到区域内切沟侵蚀发生的内在规律,其预测能力可稳定地复现于不同年份的特大降雨情况下的切沟敏感性,能够为水土流失的精准防治提供有力的科学支撑。

综上所述,通过2次独立特大降雨年的交叉验证,充分证实随机森林模型在切沟敏感性评价中不仅具备高精度,更兼具强大的稳健性和泛化能力。该模型能够真实、可靠地预测切沟在时空上的动态演化,特别是在极端降雨背景下的高发位点。

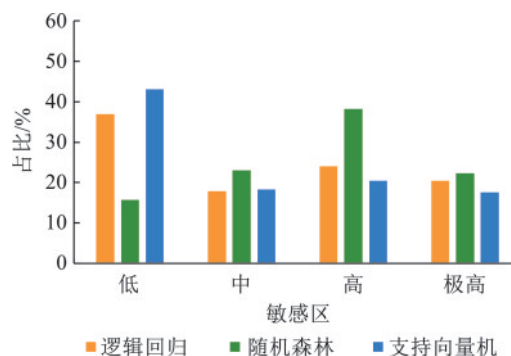


图9 2021年相对于2019年的新增切沟在各等级敏感区域内的占比

Fig.9 Proportional distribution of new gullies (2021 vs. 2019) across different gully erosion susceptibility classes

3 讨论

3.1 切沟侵蚀敏感性影响因子及其主导作用分析

通过比较随机森林模型中因子的重要性排名发现,集水区面积、距河流距离与河网密度是对模型贡献前3的变量,表明3种水文因子是影响切沟侵蚀敏感性的关键驱动因素,验证本研究的科学性与区域适应性。

集水区面积是最重要的影响因子,可以反映地表产流和汇流的空间组织特征。大集水区通常伴随着更大的汇水量和更高的侵蚀能级,特别是当坡度较大时,极易诱发下游切沟的发育^[27]。有研究^[16]表明,集水区的形状和边界条件显著影响切沟发生的空间分布特征。距河流距离被确定为第2重要的因子,此发现印证了众多关于河流对切沟侵蚀发生显著控制作用的观点。从物理机制上剖析,地表径流在汇入河道前,沿着微地形的自然沟槽汇集,随着距河流距离的缩短,汇流面积和径流量持续增大,加之水流势能的累积,最终形成强大的侵蚀力,极易诱发切沟的形成。流域中近河区域因坡度集中、水动力强,常成为切沟发生的高风险带^[22]。因此,靠近河岸的区域由于汇流效应的增强,其切沟侵蚀风险通常远高于远离河岸的区域^[22]。而河网密度作为反映水流路径复杂性与集中程度的重要指标,其高值区域通常伴随着更强的汇流能力与侵蚀强度,容易导致切沟频繁发生^[28]。高密度的河网结构意味着更高频率的汇水过程,进而促进地表能量的传递与侵蚀过程的快速响应。纳入模型中的人类活动因子在单因子排序中并不突出(图5),表明单独的人类活动因子并非切沟发生的决定性因素。值得注意的是,尽管如海拔、剖面曲率等因子在本研究中未表现出显著的重要性,但在其他区域却被视为关键因子。例如,Rahmati等^[22]研究发现,海拔在卡什干-波尔多赫塔尔流域是控制切沟形成的主导因子。在不同地貌背景下,地形因子对侵蚀过程的控制机制存在显著差异,与地质结构、土壤类型和降雨格局等多种环境变量的耦合关系密切相关。

因此,应该结合具体区域特征开展因子适配性分析,以提高模型的泛化能力。

与已有研究不同,本研究将人类活动因子(如距道路距离、距林带距离和耕作方式)纳入模型构建,弥补了以往研究多集中在自然因素的不足。交互因子重要性排序结果显示,与人类活动相关的交互因子(如耕作方式 $\times$ 距河流距离、距道路距离 $\times$ 距河流距离和距林带距离 $\times$ 距河流距离)的重要性显著高于对应单一因子,清晰地揭示人类活动(距道路距离、距林带距离和耕作方式)对切沟的影响往往不是孤立存在的,而是通过与自然环境的协同作用被放大。人类对土地的利用和扰动已经深刻地改变鹤北流域的水文地貌过程,并成为加剧土壤侵蚀的重要驱动力。在切沟敏感性评估中,必须重视“人为-自然”耦合系统的综合分析,单一因子的独立评估可能严重低估实际风险。因此,将人类活动因子的交互效应纳入模型是理解切沟侵蚀复杂性的重要突破,也为防治措施的精准化提供科学依据。未来的研究应进一步量化不同人类活动类型对切沟侵蚀风险的具体贡献^[15],为制定针对性强的土地管理策略提供科学依据。

3.2 模型性能比较与稳健性测试

本研究比较3种模型在切沟侵蚀敏感性预测中的表现,结果显示,随机森林模型(RF)实现最优的模型性能,AUC高达0.96。RF模型作为一种集成学习方法,能够有效处理高维、非线性与多重共线性问题,且无须对输入变量进行归一化处理,具备强大的数据适应能力与解释能力。与之相比,逻辑回归模型基于线性假设,难以刻画变量间的非线性交互关系,导致其在处理复杂地貌过程时表现受限。支持向量机模型虽然具备较强的拟合能力,但模型性能高度依赖于核函数选择与参数调优,在数据量较小或变量复杂时,容易出现过拟合或欠拟合的情况。其次,2019年与2021年2次特大降雨年的后验验证结果,充分彰显随机森林模型不仅在空间上实现精准预测,在时间尺度上也能表现出稳健的泛化能力,为模型的实际应用提供关键支撑。此现象并非偶然,而是由于随机森林模型通过集成多棵决策树,有效降低单一模型的不确定性,能够捕捉因子间的复杂的动态响应关系,而非仅适用于特定年份的降雨情景。因此,从模型稳定性和实用性的角度出发,RF模型在切沟侵蚀敏感性研究中具有更好的应用前景^[20]。

4 结论与展望

本研究采用随机森林(RF)、支持向量机(SVM)和逻辑回归(LR)3种机器学习模型,以鹤北流域为研究区,综合地形、水文、环境等15个因子,对切沟侵蚀敏感性进行预测与评估。本研究发现,该区域切沟

侵蚀主要受到距河流距离、集水区面积和河网密度等水文和地形因子的影响,而人类活动因子的重要性同样不容忽视。在模型性能方面,随机森林模型的预测精度显著优于其他2种模型,表明其在黑土区切沟侵蚀敏感性预测中具有最优性能和应用潜力,研究成果可为东北黑土区类似流域的切沟侵蚀防治规划提供决策支持与预警参考。

尽管本研究取得较好的结果,但仍存在一定的局限性。首先,模型构建与验证仅仅局限于2个相邻流域,未来需要考虑更多不同地貌、气候和耕作背景的流域数据,以检验模型的普适性;其次,所选因子是静态的,未能反映年内的动态变化;最后,10 m分辨率的DEM数据在刻画微地貌特征(如垄沟、田埂)方面仍显不足。未来的研究应着力于融合更高分辨率的无人机遥感影像,精细刻画微地貌与人类活动因子;构建动态模型,耦合年内降雨过程、作物生长和耕作措施,实现从“切沟空间敏感性”向“切沟时空风险预警”的转变。

参考文献:

- [1] 姜芸,王军,张莉.东北典型黑土区侵蚀沟形态及分布特征[J].农业工程学报,2020,36(7):157-165.
Jiang Yun, Wang Jun, Zhang Li. Morphology and distribution characteristics of erosion gully in the typical black soil region of northeast China[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(7): 157-165.
- [2] 王文娟,邓荣鑫,张树文,等.东北典型黑土区侵蚀沟发育演化特征分析[J].地理与地理信息科学,2019,35(4):105-110.
Wang Wenjuan, Deng Rongxin, Zhang Shuwen, et al. Development evolution characteristics of gully erosion in typical black soil region of northeast China[J]. Geography and Geo-Information Science, 2019, 35(4): 105-110.
- [3] Poesen J, Nachtergaele J, Verstraeten G, et al. Gully erosion and environmental change: Importance and research needs[J]. Catena, 2003, 50(2/3/4): 91-133.
- [4] Gudino-Elizondo N, Biggs T, Bingner R, et al. Modeling ephemeral gully erosion from unpaved urban roads: Equifinality and implications for scenario analysis[J]. Geosciences, 2018, 8(4): e137.
- [5] Rodriguez-Caballero E, Cantón Y, Chamizo S, et al. Effects of biological soil crusts on surface roughness and implications for runoff and erosion[J]. Geomorphology, 2012, 145/146: 81-89.
- [6] Chaplot V, Brown J, Dlamini P, et al. Rainfall simulation to identify the storm-scale mechanisms of gully bank retreat [J]. Agricultural Water Management, 2011, 98(11): 1704-1710.
- [7] Thomas J T, Iverson N R, Burkart M R, et al. Long-term

- growth of a valley-bottom gully, western Iowa [J]. *Earth Surface Processes and Landforms*, 2004, 29(8): 995-1009.
- [8] Arabameri A, Rezaei K, Pourghasemi H R, et al. GIS-based gully erosion susceptibility mapping: A comparison among three data-driven models and AHP knowledge-based technique[J]. *Environmental Earth Sciences*, 2018, 77(17): e628.
- [9] Pourghasemi H R, Yousefi S, Kornejady A, et al. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling[J]. *Science of the Total Environment*, 2017, 609: 764-775.
- [10] Huang Donghao, Su Lin, Fan Haoming, et al. Identification of topographic factors for gully erosion susceptibility and their spatial modelling using machine learning in the black soil region of northeast China[J]. *Ecological Indicators*, 2022, 143: e109376.
- [11] Saha S, Sarkar R, Thapa G, et al. Modeling gully erosion susceptibility in Phuentsholing, Bhutan using deep learning and basic machine learning algorithms[J]. *Environmental Earth Sciences*, 2021, 80(8): e295.
- [12] 李珊珊, 王春梅, 唐杰, 等. 东北黑土区典型小流域浅沟向切沟发育及其地形临界研究[J]. *武汉大学学报(信息科学版)*, 2026, 51(2): 345-356.
- Li Shanshan, Wang Chunmei, Tang Jie, et al. Development processes from ephemeral to permanent gullies and their critical terrain thresholds in a typical watershed of the mollisol region, northeast China[J]. *Geomatics and Information Science of Wuhan University*, 2026, 51(2): 345-356.
- [13] Tang Jie, Liu Gang, Xie Yun, et al. Ephemeral gullies caused by snowmelt: A ten-year study in northeastern China[J]. *Soil and Tillage Research*, 2021, 212: e105048.
- [14] Gayen A, Pourghasemi H R, Saha S, et al. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms[J]. *Science of the Total Environment*, 2019, 668: 124-138.
- [15] Nagarajan R, Roy A, Vinod Kumar R, et al. Landslide hazard susceptibility mapping based on terrain and climatic factors for tropical monsoon regions[J]. *Bulletin of Engineering Geology and the Environment*, 2000, 58(4): 275-287.
- [16] Conforti M, Aucelli P P C, Robustelli G, et al. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (northern Calabria, Italy)[J]. *Natural Hazards*, 2011, 56(3): 881-898.
- [17] Arabameri A, Pradhan B, Lombardo L. Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling[J]. *Catena*, 2019, 183: e104223.
- [18] Garosi Y, Sheklabadi M, Conoscenti C, et al. Assessing the performance of GIS- based machine learning models with different accuracy measures for determining susceptibility to gully erosion[J]. *Science of the Total Environment*, 2019, 664: 1117-1132.
- [19] Moore I D, Burch G J, MacKenzie D H. Topographic effects on the distribution of surface soil water and the location of ephemeral gullies [J]. *Transactions of the ASAE*, 1988, 31(4): 1098-1107.
- [20] Kumari M, Singh C K, Bakimchandra O, et al. Geographically weighted regression based quantification of rainfall-topography relationship and rainfall gradient in central Himalayas [J]. *International Journal of Climatology*, 2017, 37(3): 1299-1309.
- [21] Freeman T G. Calculating catchment area with divergent flow based on a regular grid[J]. *Computers and Geosciences*, 1991, 17(3): 413-422.
- [22] Rahmati O, Tahmasebipour N, Haghizadeh A, et al. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion[J]. *Geomorphology*, 2017, 298: 118-137.
- [23] Tang Jie, Liu Gang, Xie Yun, et al. Annual variation of ephemeral gully erosion in a cultivated catchment [J]. *Geoderma*, 2021, 401: e115166.
- [24] Tang Jie, Liu Gang, Xie Yun, et al. Effect of topographic variations and tillage methods on gully erosion in the black soil region: A case-study from northeast China [J]. *Land Degradation and Development*, 2022, 33(18): 3786-3800.
- [25] Conoscenti C, Sheshukov A Y. Regional variability of terrain index and machine learning model applications for prediction of ephemeral gullies [J]. *Geomorphology*, 2023, 442: e108915.
- [26] Akgün A, Türk N. Mapping erosion susceptibility by a multivariate statistical method: A case study from the Ayvalık region, NW Turkey[J]. *Computers and Geosciences*, 2011, 37(9): 1515-1524.
- [27] Nefeslioglu H A, Gokceoglu C, Sonmez H. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps [J]. *Engineering Geology*, 2008, 97(3/4): 171-191.
- [28] 温艳茹, 余强毅, 杨扬, 等. 黑土小流域沟道分布遥感监测及主控因素研究[J]. *农业资源与环境学报*, 2021, 38(6): 1074-1083.
- Wen Yanru, Yu Qiangyi, Yang Yang, et al. Gully distribution monitoring based remote sensing and key driving factors in the catchment scale of mollisol region[J]. *Journal of Agricultural Resources and Environment*, 2021, 38(6): 1074-1083.