

DOI:10.13869/j.cnki.rswc.2026.01.040; CSTR:32311.14.rswc.2026.01.040.

田向阳, 杜鹏飞, 赵莹, 等. 基于随机森林模型的黄土高原典型小流域不同样本密度的沟蚀风险预测[J]. 水土保持研究, 2026, 33(1):101-110.

Tian Xiangyang, Du Pengfei, Zhao Ying, et al. Prediction of gully erosion risk at different sample densities in typical small watersheds of Loess Plateau based on random forest model[J]. Research of Soil and Water Conservation, 2026, 33(1): 101-110.

基于随机森林模型的黄土高原典型小流域 不同样本密度的沟蚀风险预测

田向阳, 杜鹏飞, 赵莹, 陈吟

(中国水利水电科学研究院 国际泥沙研究培训中心, 北京 100048)

摘要: [目的] 探讨黄土高原典型小流域沟蚀风险预测的最佳样本密度, 并分析影响沟蚀发生的主要特征因子, 为侵蚀沟的防治提供参考依据。[方法] 采用随机森林模型结合 SHAP 算法, 分别选取 25%, 50%, 75% 和 100% 的侵蚀沟样本密度进行沟蚀风险预测并量化主导因子对模型输出的贡献。[结果] 50% 样本密度风险预测的效果最佳, 准确率、精确率、Kappa C 值分别达到 0.901, 0.894, 0.802, 明显优于 25% (0.871, 0.851, 0.743), 且高于 75% (0.898, 0.882, 0.795) 和 100% (0.899, 0.880, 0.798)。AUC 值依次为 0.924, 0.956, 0.956, 0.959, 召回率依次为 0.887, 0.910, 0.917, 0.924, 50% 样本密度的 AUC 值和召回率与 75% 和 100% 差距微小, 因此在保证准确的前提下认为 50% 样本密度是最优选择。[结论] 在各种影响因素中, 土地利用类型对沟蚀风险的预测贡献最大, 平面曲率和坡度次之。沟蚀风险与侵蚀严重程度耦合分析表明低风险区域也可能发生中度和重度侵蚀, 说明沟蚀风险高低不能代表侵蚀严重程度, 可为黄土高原类似条件下的流域沟蚀评估和防治提供参考依据。

关键词: 沟蚀风险预测; 随机森林; 黄土高原; SHAP 算法

中图分类号: S157

文献标识码: A

文章编号: 1005-3409(2026)01-0101-10

Prediction of gully erosion risk at different sample densities in typical small watersheds of Loess Plateau based on random forest model

Tian Xiangyang, Du Pengfei, Zhao Ying, Chen Yin

(International Research and Training Center on Erosion and Sedimentation,
China Institute of Water Resources and Hydropower Research, Beijing 100048, China)

Abstract: [Objective] The study explores the optimal sample density for gully erosion risk prediction in typical small watersheds of the Loess Plateau and analyzes the main feature factors influencing gully occurrence, thereby providing a reference for gully erosion prevention and control. [Methods] The Random Forest (RF) model combined with SHAP algorithm was used to predict gully erosion risks at sample densities of 25%, 50%, 75%, and 100% of the total gully samples, and the contributions of the dominant factors to the model output were quantified. [Results] The 50% sample density achieved the best predictive performance, with accuracy, precision, and Kappa coefficient reaching 0.901, 0.894, and 0.802 respectively. These values significantly exceeded those of the 25% density (0.871, 0.851, 0.743), and were higher than the 75% (0.898, 0.882, 0.795) and 100% (0.899, 0.880, 0.798) density. The AUC values were 0.924, 0.956, 0.956, and 0.959 respectively across the four densities, and the recall rates were 0.887, 0.910, 0.917, and 0.924 respectively. Notably, the 50% density showed negligible differences in AUC value and recall rate compared to the 75% and 100% densities, so it was considered the optimal choice under the premise of ensuring accuracy. [Conclusion] Among all influence

收稿日期: 2025-03-07

修回日期: 2025-04-18

接受日期: 2025-06-01

资助项目: 国家自然科学基金项目(U2243213); 国家重点研发计划“政府间国际科技创新合作”项目(2021YFE0113800); 中国水利水电科学研究院青年托举项目(SC110145B0012023)

第一作者: 田向阳(2000—), 男, 河北保定人, 硕士研究生, 研究方向为土壤侵蚀和水土保持研究。E-mail: hbndtxy@163.com

通信作者: 杜鹏飞(1982—), 男, 内蒙古赤峰人, 博士, 正高级工程师, 主要从事土壤侵蚀和水土保持研究。E-mail: dupf@iwhr.com

<http://stbcyj.paperonice.org>

factors, land use type contributes the most to the gully erosion risk prediction, followed by slope and planar curvature. The coupling analysis of gully erosion risk and erosion severity indicates that moderate and severe erosion may occur in low-risk areas, suggesting that the gully erosion risk levels cannot fully represent the erosion severity, which provides a reference for gully erosion assessment and prevention in watersheds with similar conditions across the Loess Plateau.

Keywords: gully erosion risk prediction; random forest; Loess Plateau; SHAP algorithm

黄土高原地形破碎、土质疏松,在暴雨影响下水土流失严重,对农牧业的发展造成了严重的不利影响^[1]。黄土高原主要的侵蚀类型为水力侵蚀,沟道作为黄土高原流域系统中最活跃的部分,分布广泛、形态复杂^[2]。研究表明,极端暴雨条件下,黄土高原坡耕地侵蚀最为严重且以沟蚀为主,使该区域成为沟蚀最严重的地区^[3]。侵蚀沟的发育不仅加剧流域水土流失,还破坏土地资源,严重威胁当地生态环境与社会发展^[4],因此加强侵蚀沟的防治至关重要,而沟蚀风险预测可为水土保持工程和措施的实施提供科学依据^[5],成为黄土高原最近研究的热点问题之一。

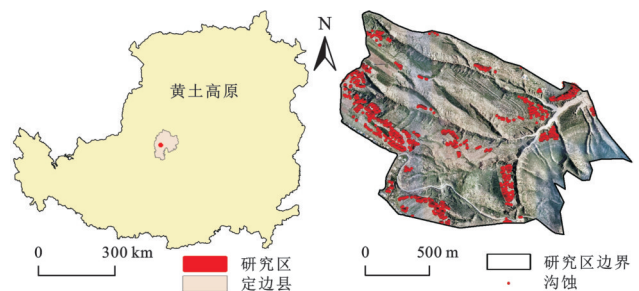
近年来,包括随机森林(Random Forest, RF)、极限梯度提升(XGBoost)、轻量梯度提升机(LightGBM)等在内的机器学习预测模型已逐渐成为沟蚀风险评估的重要工具。已有研究表明,在预测精度上,RF模型在不同研究区域均表现出色^[6-7],相较之下,LightGBM和XGBoost仅在部分研究中表现较佳^[8-9]。在适用性方面,RF模型在沟蚀风险预测中同样表现优越,并在不同地区的小流域得到了广泛应用^[10-11]。在特征因子方面,地形因子(高程、坡度等)和土地利用类型是影响沟蚀最重要且最常用的因子。已有研究发现,高程、坡度在多个地区均为主导因子^[6-7],而一些地区土地利用的影响更为显著^[12]。在样本密度对模型预测的影响方面,曹文悦^[8]在云南元江干热河谷地区利用RF等多种机器学习模型分析了沟蚀易发性,结果表明当抽样密度为25%时预测效果最好。而针对黄土高原的类似研究不多见,虽然已有研究成功构建了基于机器学习的沟头侵蚀模型^[7],并探索了特征因子在沟蚀风险评估中的作用,但这些研究大多基于常规降雨,未考虑极端暴雨事件对沟蚀形成的影响。近几十年来黄土高原地区暴雨事件呈显著增强趋势,降水强度、频率和极端性不断上升^[13-14],频发的极端暴雨事件极易诱发侵蚀沟的大量形成与迅速扩展^[15-16],因此,开展不同样本密度对沟蚀风险预测的影响研究具有重要的现实意义。

为此,本研究以2022年7月10日定边县极端暴雨事件为研究背景,采用实地调查、无人机航拍、影像解译等方法,获取黄土高原典型小流域的沟蚀数据,应

用RF模型进行风险预测。为探讨样本密度对预测精度的影响,设置25%,50%,75%和100%共4种不同样本密度进行对比分析,旨在提升风险预测的准确性和效率,为黄土高原地区的沟蚀防治工作提供参考。

1 研究区概况

研究区位于黄土高原中部的晋陕蒙丘陵沟壑区,地形起伏大、沟壑纵横,在行政区域上隶属于陕西省榆林市定边县,属于陕北盖沙丘陵沟壑拦沙防沙区,受风沙与水蚀的双重作用影响,水土流失严重。气候类型上属于温带半干旱季风气候,春季多风沙,夏季干旱少雨,年平均气温为7.9℃,多年平均降雨量为316.9mm,降雨量年际变化较大,主要集中在7—9月^[17]。研究小流域位于定边县的中西部(如图1所示),土壤疏松、抗蚀能力弱,沟蚀现象典型且严重,其面积约为2.35km²,海拔为1622~1833m,土地利用类型以草地、沟道和耕地为主,此外还包括建设用地、道路和水域。2022年7月10—11日,定边县发生极端降雨,暴雨中心累积最大降雨量为114.3mm,累计降雨量达91.1mm,1h最大雨强为56.3mm,单次降雨超全年总降雨量的1/4^[18]。



注:基于标准地图服务系统下载的标准图号GS(2019)1822号的标准地图制作,底图未做修改,下图同。

图1 研究区位置与沟蚀样点分布

Fig. 1 Location of study area and distribution of gully erosion sample points

2 数据获取及研究方法

2.1 数据获取

2.1.1 侵蚀沟数据获取 暴雨发生后,由中国水利水电科学研究院和中国科学院水利部水土保持研究所联合组成的调查小组即赴定边县开展暴雨水土流

失调查。通过大疆精灵 4RTK 无人机航摄获得研究区的遥感影像(分辨率为 0.07 m),获取研究区土地利用情况、侵蚀沟尺寸(包括沟深、沟宽、沟长等数据)以及农作物等资料。

调查结束后,通过目视解译影像,确定研究小流域共发育 1 659 条侵蚀沟。根据侵蚀沟长度,将其分成 4 个级别,分别为: ≤ 10 m, 10 m < 长度 ≤ 20 m, 20 m < 长度 ≤ 30 m, > 30 m, 4 个级别分别均匀地标记 1—4 个侵蚀点,如表 1 所示。在参考相关研究^[19]的基础上,结合研究小流域实际情况,根据沟宽将侵蚀沟分成 3 个级别: ≤ 0.2 m、 0.2 m < 宽度 ≤ 0.5 m、 > 0.5 m, 3 种不同的沟宽分别记为 1、2、3 个等级,如表 2 所示。最终共标记侵蚀点 2 589 个,并通过 ArcGIS 生成同样数量的非侵蚀点,经检查确保非侵蚀点未落在侵蚀沟上,此为 100% 样本密度,在此基础上,利用随机采样的方法分别抽取 25%、50% 和 75% 的侵蚀点和非侵蚀点样本,作为模型预测的不同数据基础。

2.1.2 特征因子的选取 侵蚀沟的发生是自然因素和人为因素共同作用的结果,自然因素主要包括地形地貌、土壤条件、气候条件等,人为因素主要包括植被破坏和不合理耕作,表现为土地利用类型的不合理分配^[20]。本研究选择土地利用类型及 13 个和地

形相关的因子,包括高程、剖面曲率、平面曲率、坡度、坡向、坡长、汇水面积、距河流距离、地形粗糙度、高程变异系数、地表切割深度、地形起伏度、距道路距离,共计 14 个特征因子进行分析^[21]。

表 1 侵蚀沟长度的分级标准

Table 1 Grading standards of gully length

侵蚀沟长度/m	各长度标记的沟蚀点数量	侵蚀沟数量	各长度对应的沟蚀点数量
≤ 10 m	1	1093	1093
(10 m, 20 m]	2	332	664
(20 m, 30 m]	3	104	312
> 30 m	4	130	520
合计	—	1659	2589

表 2 侵蚀沟宽度的分级标准

Table 2 Grading standards of gully width

侵蚀沟宽度/m	侵蚀等级
≤ 0.2 m	1
(0.2 m, 0.5 m]	2
> 0.5 m	3

其中,土地利用类型基于遥感影像,采用人工目视解译的方法进行分类;其余 13 个地形因子基于 DEM (0.30 m 分辨率),利用 ArcGIS 中的空间分析等工具逐一提取。上述特征因子的提取结果如图 2 所示。

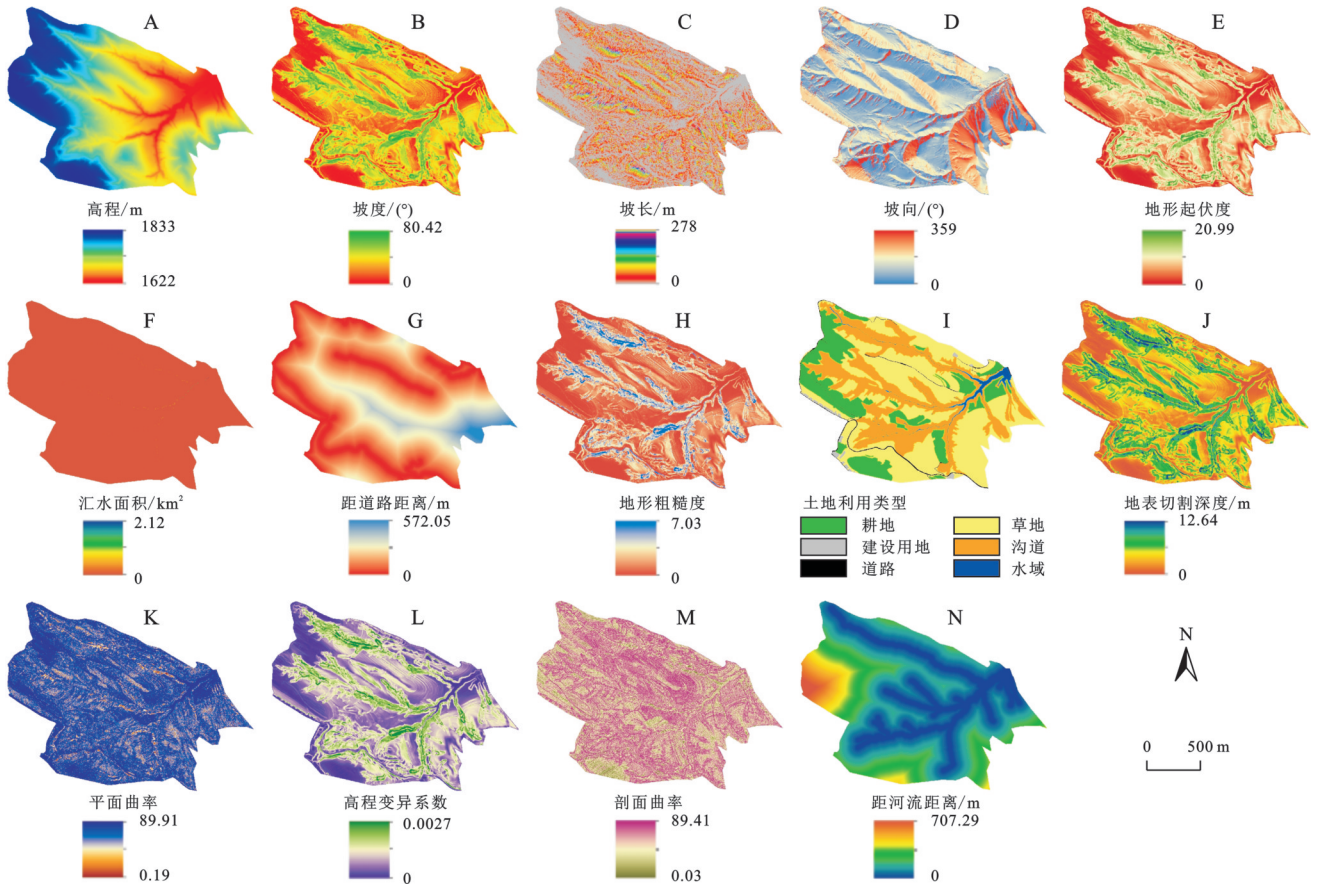


图 2 特征因子图

Fig. 2 Feature factor maps

2.2 研究方法

2.2.1 特征因子相关性检验 在采用机器学习模型时,如果特征因子之间存在相关性,一定程度上会降低模型预测结果的准确性,并干扰因子的重要性排序^[22]。为避免冗余数据对模型的影响,首先要对因子进行相关性检验,本研究采用皮尔逊相关系数分析方法对初步选取的土地利用类型、高程、距河流距离等 14 个因子进行相关性分析。

两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商:

$$\rho_{X,Y} = \frac{\text{COV}(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y} \quad (1)$$

式中: $\rho_{X,Y}$ 为整体相关系数; X, Y 为两个变量的值; $\text{COV}(X, Y)$ 是协方差; σ_X 和 σ_Y 分别是 X 和 Y 的标准差。

皮尔逊相关系数是衡量两个变量之间线性关系强度和方向的统计指标。它的范围从-1到1,当两个因子间的相关系数小于0.3时,说明因子相关性较差,位于0.3~0.5时,呈现低相关,位于0.5~0.8时,说明中度相关,当 ≥ 0.8 时,被认为是高度相关^[23]。

2.2.2 随机森林模型 随机森林模型是目前十分热门的机器学习算法之一,通过构建多棵决策树并综合其预测结果来实现分类和回归任务。该模型能够降低共线性对模型性能的影响,对缺失和非平衡的数据有较好的容忍度,且无需对数据进行归一化处理,可以提高模型的准确性和鲁棒性评价效果优良^[24]。

本研究应用随机森林模型进行沟蚀风险预测,模型训练采用 Python 语言,并在 PyCharm 环境下完成,应用重复随机划分交叉验证的方法进行训练。设置参数时,采用 200 棵决策树,最大深度为 30,分别使用侵蚀点和等量非侵蚀点样本的 70% 作为训练集对模型进行训练,剩余 30% 作为测试集预测沟蚀风险分布,将风险预测结果通过自然断点法分为非常低、低、中、高、非常高风险 5 个等级。根据评价指标选出最好的样本密度后应用随机森林模型按照侵蚀沟宽度划分的侵蚀等级(表 2),非侵蚀点为无侵蚀,等级 1 为轻度侵蚀,等级 2 为中度侵蚀,等级 3 为重度侵蚀,对侵蚀严重程度进行预测。

2.2.3 精度评价 准确率(Accuracy)、召回率(Recall)、精确率(Precision)、AUC 值和 Kappa C 值是用来评估机器学习模型性能的常见指标^[25]。本研究采用上述 5 个指标分析计算 25%、50%、75% 和 100% 共 4 种不同样本密度在随机森林模型下的精度。5 个评价指标通过混淆矩阵获取的数据计算,具体计算方法如下:

准确率(Accuracy):正确分类沟蚀和非沟蚀样本

数占测试集总样本数的比例,表征算法的分类精度。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

式中:TP 为测试集中实际为沟蚀且被正确预测为沟蚀的样本个数;TN 为测试集中实际为非沟蚀且被正确预测为非沟蚀的样本个数;FP 为测试集中实际为非沟蚀但被错误预测为沟蚀的样本个数;FN 为测试集中实际为沟蚀但被错误预测为非沟蚀的样本个数;FP、FN 为错误分类的样本个数;TP、TN 为正确分类的样本个数。

精确率(Precision):预测为沟蚀的样本中,真实为沟蚀的样本个数所占的比例,衡量预测结果的可靠性。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

召回率(Recall):在真实为沟蚀的样本中,被正确预测为沟蚀样本个数所占的比例,衡量模型对沟蚀的识别能力。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

ROC 曲线表征敏感性和特异性的相互关系,AUC 值是 ROC 曲线下与坐标轴围成的面积^[26],其中模型的特异性(假阳性率)为 x 轴,敏感度(真阳性率)为 y 轴。AUC 值的范围一般为 0.5~1.0,值越高,模型预测性能越好。

Kappa C 值是基于混淆矩阵的检验一致性和分类准确性的指标,该系数同时考虑了观察一致性和偶然一致性。

$$\text{Kappa C} = \frac{p_{\text{obs}} - p_{\text{exp}}}{1 - p_{\text{exp}}} \quad (5)$$

$$p_{\text{obs}} = \frac{\text{TP} + \text{TN}}{N_c} \quad (6)$$

$$p_{\text{exp}} = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{FP} + \text{TN})(\text{TN} + \text{FN})}{N_c^2} \quad (7)$$

式中: p_{obs} 是观测到的一致性比例; p_{exp} 是预期的一致性比例; N_c 是地图中单元格的总数量。

2.2.4 SHAP 算法 对于随机森林模型而言,虽然能够获取特征因子的重要性排序,但这并不足以确保每个特征因子值的贡献是公平的,也难以清晰地了解各个特征因子是如何影响预测结果的。SHAP 算法作为一种基于博弈论中 Shapley 值的机器学习解释性工具,能够确保每个特征因子值的贡献是公平的,并有效避免因子之间共线性可能导致的偏误,不仅提供全局解释,揭示整个数据集上特征因子的重要性分布,还具备局部解释能力,能对特定的观测

值深入分析每个因子如何影响风险预测^[27]。因此,在本研究采用 SHAP 值量化每个因子对沟蚀风险预测结果的影响。

对于特征集合 S 中的特征 i , Shapley 值计算公式如下:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [\nu(S \cup \{i\}) - \nu(S)] \quad (8)$$

式中: Φ_i 为特征 i 的 SHAP 值(即 Shapley 值); N 表示所有特征的集合; $|N|$ 是总特征数量; $S \subseteq N \setminus \{i\}$ 是不包含特征 i 的任何一个特征子集; $|S|$ 是集合 S 中的特征数量; $\nu(S)$ 是特征子集 S 对模型预测输出的贡献; $\nu(S \cup \{i\})$ 是包含特征 i 的特征集对模型预测输出的贡献。

3 结果与分析

3.1 特征因子相关性分析

本研究中利用 Python 处理 14 个特征因子,从中提取出 $10 \text{ m} \times 10 \text{ m}$ 分辨率下的原始值并整理成表格,再通过 Origin 软件对这些原始值进行可视化处

理,绘制出皮尔逊相关性热图,以直观展示特征因子之间的相关性,如果两个因子的相关性系数大于 0.5,表明二者具有很高的相关性,不能同时出现。

图 3 绘制了特征因子之间的相关性热图,结果表明坡度与地形起伏度、地形粗糙度、地表切割深度、高程变异系数相关性系数均大于 0.5,说明这些因子可能表征相似的地貌信息,坡度是导致这几个因子的最直接原因,也是侵蚀沟之所以发生的最基础条件之一,同时还是土壤侵蚀相关模型中最为重要的考量因素之一,为此本研究保留坡度因子。此外,高程与距河流距离相关性系数为 0.65,二者的区别在于,高程重点反映的是地形对侵蚀沟的塑造作用,而距河流距离重点强调的水文过程上的联系。相较之下,高程更能通过地貌特征和径流汇集全面反映侵蚀沟发育的基础条件和发展过程,所以保留高程因子。本研究最终选取的沟蚀风险预测的评价指标包括高程、坡度、坡长、坡向、汇水面积、距道路距离、土地利用类型、平面曲率和剖面曲率 9 个因子。

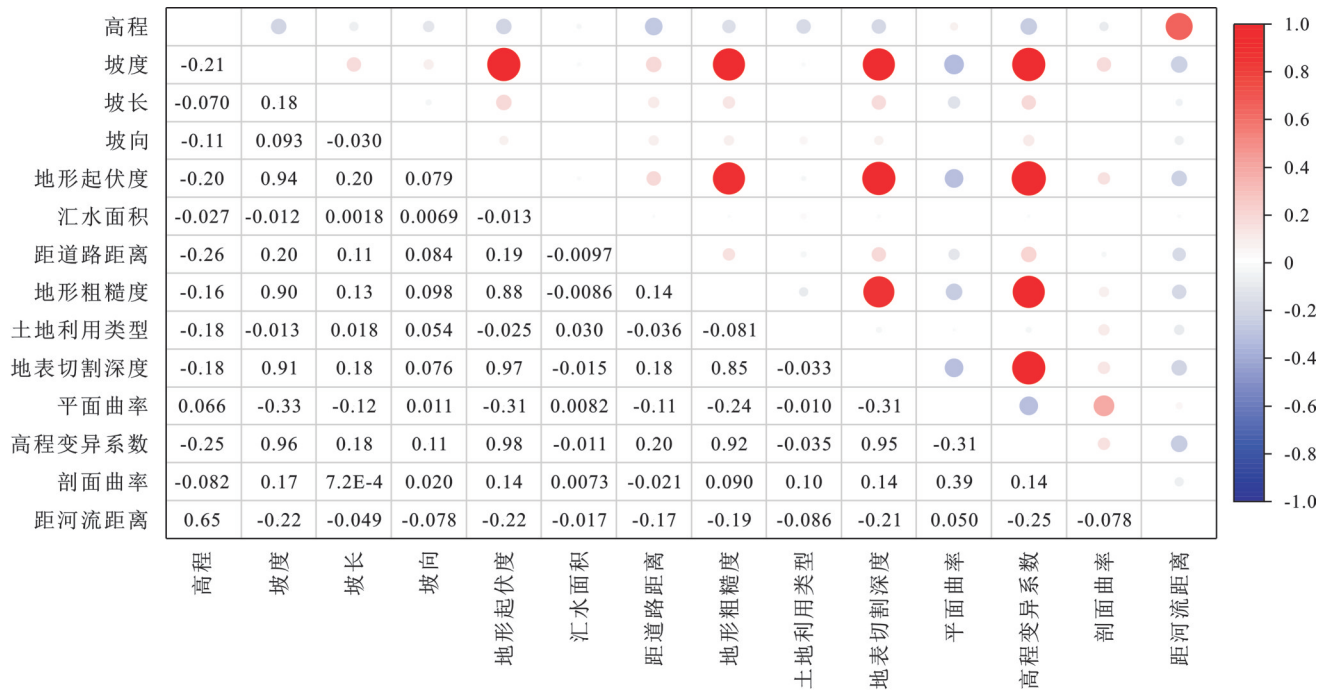


图 3 皮尔逊相关性热图

Fig. 3 Pearson correlation heatmap

3.2 模型预测性能分析

表 3 和图 4 给出了在 4 种不同样本密度随机森林模型精度计算的结果和 ROC 曲线图。从准确率指标来看,50% 样本密度下指标值最高为 0.901,25% 样本密度下最低为 0.871,说明 50% 样本密度下模型分类效果是最好的。从召回率指标来看,50% 样本密度下的指标值为 0.910,虽略小于 100% 样本密度(0.924),但与 75%(0.917)差距很小,且明显高于 25%

(0.887)。从精确率指标来看,50% 样本密度下指标值最高为 0.894,高于 75%(0.882)和 100%(0.880),而 25% 样本密度下精确率最低为 0.851;说明 50% 样本密度下模型预测为正类的样本中实际为正类的比例最高。从 AUC 指标来看,如图 4 所示,ROC 曲线中曲线越接近左上角,即 AUC 值越大,预测能力越强。50% 样本密度的指标值为 0.956,虽略小于 100% 样本密度(0.959),但与 75%(0.956)一样,均明显高于

25% 样本密度下的值(0.924)。从 Kappa C 值来看, 50% 样本密度的指标值最高为 0.802, 说明该样本密度下模型的检验一致性和分类准确性最好。

表 3 不同样本密度下模型精度评价表

Table 3 Model accuracy evaluation under different sample densities

项目	样本密度			
	25%	50%	75%	100%
准确率	0.871	0.901	0.898	0.899
召回率	0.887	0.910	0.917	0.924
精确率	0.851	0.894	0.882	0.880
AUC 值	0.924	0.956	0.956	0.959
Kappa C	0.743	0.802	0.795	0.798

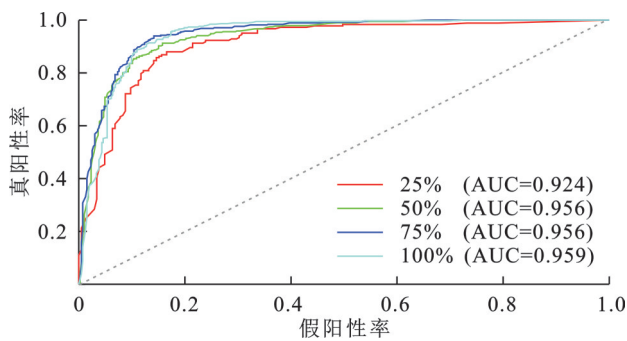


图 4 ROC 曲线图
Fig. 4 ROC curves

综上, 通过上述评价指标的对比分析可以发现, 不同样本密度在随机森林模型下都有良好的表现, 但综合来看, 50% 样本密度下准确率、精确率和 Kappa C 值都是最高的, 召回率和 AUC 值虽然小于

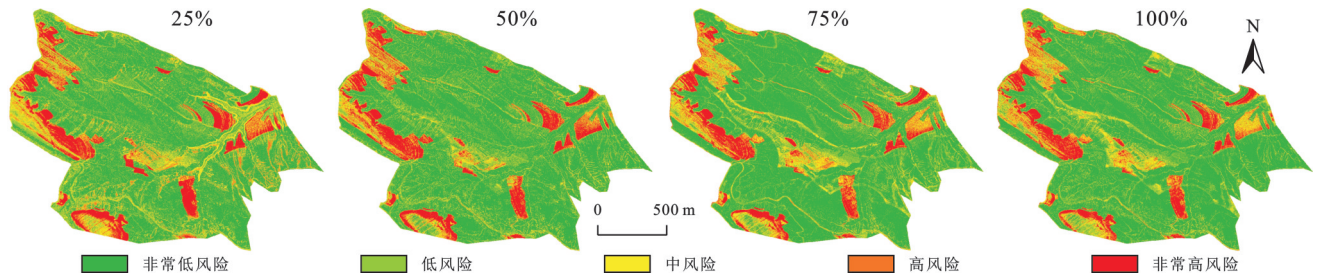


图 5 不同样本密度下沟蚀风险预测图

Fig. 5 Gully erosion risk prediction maps under different sample densities

表 4 不同样本密度下风险分布表

Table 4 Risk distribution under different sample densities %

风险程度	样本密度			
	25%	50%	75%	100%
非常低风险	54.58	63.14	61.91	63.10
低风险	18.45	14.96	16.96	16.40
中风险	10.44	8.05	8.14	7.93
高风险	8.18	6.01	6.18	5.88
非常高风险	8.35	7.84	6.82	6.69

100% 样本密度, 但相差仅有 0.014 和 0.003, 所以认为当样本密度为 50% 时, 对研究区沟蚀风险的预测效果是最准确的。

3.3 沟蚀风险及特征因子分析

3.3.1 沟蚀风险空间分布 图 5 为 25%, 50%, 75% 和 100% 样本密度下计算的沟蚀风险预测图。从图中可以看出, 低风险区域在 4 种样本密度预测结果下均占据了绝大部分面积, 反映出研究区中大多数位置发生沟蚀的风险较低, 高风险区域主要集中在地形起伏大、土地利用类型过渡的局部区域。表 4 为不同样本密度下沟蚀风险分布表。从结果来看, 随着样本密度的增加, 非常低风险区域的占比呈现轻微波动, 在 50% 样本密度下达到 63.14%, 略高于其他样本密度的结果, 低风险区域占比为 14.96%, 中风险、高风险、非常高风险区域面积占比和其他样本密度相比, 变化幅度小且较为平稳。

基于上述分析, 50% 的样本密度在侵蚀沟灾害风险预测中, 各风险等级占比较其他样本密度的预测结果更趋稳定, 综合精度评价结果, 在确保预测准确的前提下, 认为 50% 的样本密度是最优选择。然而, 此研究仍存在一定局限性。首先, 在其他区域 50% 样本密度的预测效果可能不是最优的, 未来需进一步验证其适用性。其次, 本研究评估了 25%、50%、75% 和 100% 样本密度的预测效果, 尚不清楚其他样本密度是否仍能保持较高的预测精度。因此, 未来研究可进一步探索更广泛的样本密度范围, 并结合不同研究区的特征, 以提高模型的适用性和普遍性。

3.3.2 特征因子分析 沟蚀的成因复杂, 识别主导沟蚀的特征因子对于制定科学的防治策略具有重要意义。本研究结合 SHAP 算法和随机森林模型, 对特征因子进行重要性分析。虽然传统的随机森林模型能评估因子重要性, 但无法解释各变量如何影响预测结果^[28], 而 SHAP 算法能进一步揭示各特征因子的贡献程度。图 6 展示了沟蚀风险预测中各特征因子的重要性排序, 量化了每个特征因子对预测结果的影响程度。SHAP 值越大, 表明该因子对模型预测的影响越显著。从图 6 可以看出, 土地利用类型对

沟蚀风险预测的影响最大,显著高于其他因子,表明其在模型中的主导地位,这与 Amiri 等的结论一致,其在评估伊朗 Maharloo 流域的沟蚀发生风险中,也发现土地利用是影响当地沟蚀的主要因素^[12]。平面曲率和坡度分别位列第二和第三位,表明了地形特征的重要性,其余因子对预测结果的影响相对较低。

图 7 为 50% 样本密度下侵蚀沟灾害风险预测的 SHAP 特征摘要图,用于分析不同特征因子对模型输出的影响程度及方向。点的颜色表示特征值的大小,其中红色表示高值,蓝色表示低值。由图 7 可见,

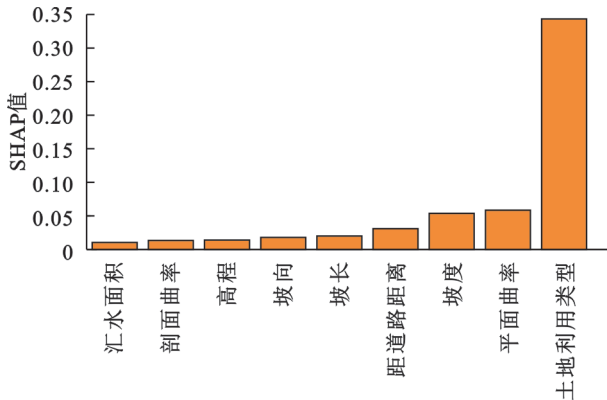


图 6 特征因子重要性排序图

Fig. 6 Importance ranking diagram of feature factors

图 8 为主导因子(土地利用类型、坡度、平面曲率、距道路距离)的单因子依赖图,可以直观揭示主导因子与沟蚀风险之间的关系。从土地利用类型的影响上看,耕地对沟蚀的发生贡献度最大,研究区土壤为疏松的黄土,在降雨作用下极易发生侵蚀。同时,由于人为耕作扰动导致的地表结构破

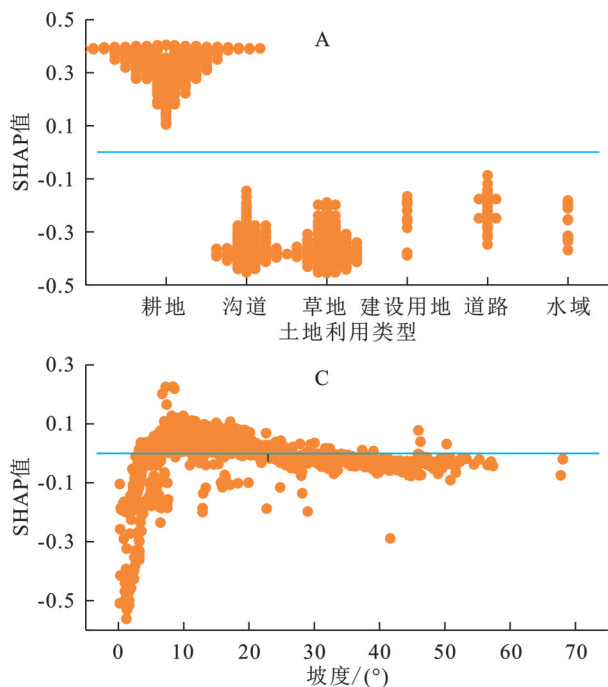


图 8 主导因子与模型预测结果的依赖关系

Fig. 8 Dependence relationships between dominant factors and model prediction results

土地利用类型的影响最为显著,其 SHAP 值分布范围广,表明高值对模型输出具有显著的负向影响,而低值则产生正向影响。坡度对沟蚀的影响也较为显著,其 SHAP 值随坡度增大先逐渐增大,随后减小至接近 0。平面曲率的 SHAP 值随特征值增大而逐渐升高,表明平面曲率越大,对风险预测的正向贡献越明显。距道路距离的高值则倾向于负向影响,说明远离道路的区域风险较低。其他特征如高程等,其 SHAP 值多集中在 0 附近,表明它们对预测结果的影响很小。

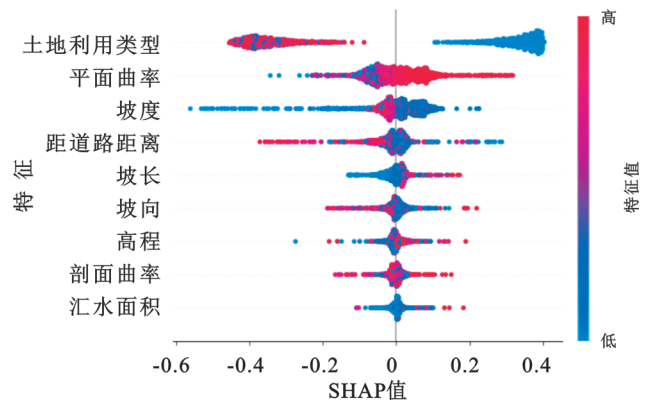


图 7 SHAP 特征摘要图

Fig. 7 SHAP feature summary plot

坏,提高了土壤侵蚀的发生风险。此外,暴雨作用下,耕地表层的垄沟容易在短时间内积聚大量汇水,显著增加了沟蚀风险。草地和沟道对沟蚀发生的负向影响较大,草地能有效降低降雨直接作用于地表的动能,同时其根系可增强土壤的稳定性,降低侵蚀沟发生率。

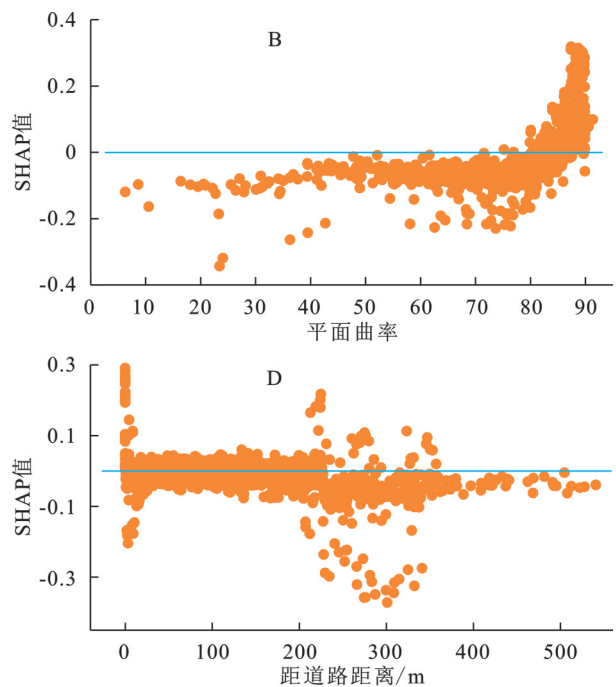


图 8 主导因子与模型预测结果的依赖关系

Fig. 8 Dependence relationships between dominant factors and model prediction results

此外,草地较强的入渗能力可减少地表径流的产生,降低水流的侵蚀作用,因此草地在一定程度上抑制了沟蚀的发生;沟道一般已经历长期侵蚀形成相对稳定的坡度和结构,沟壁上的植被或坚硬的基岩减弱了沟蚀的发生。建设用地、道路和水域则对沟蚀的影响较小,整体表现为负向作用,建设用地和道路由于硬质地表的存在,抗侵蚀能力较强,受降雨冲刷影响较小;水域不受雨水和地表径流的直接冲刷,还能起到削弱地表径流的作用,同时也无法监测到水域底部的侵蚀或沉积情况,因此这3类土地利用类型对沟蚀发生有一定的抑制作用。

从坡度的影响上看,坡度小于 5° 时对沟蚀的发生有负向影响,坡度越小,减弱沟蚀效果越明显,随着坡度增大,影响逐渐减弱;坡度 $5^{\circ}\sim 20^{\circ}$ 时对沟蚀的发生起促进作用,这与耕地的下边缘以及土地利用类型变化等原因高度相关;坡度大于 20° 时,对沟蚀的发生有很小的负向影响。此时,对应的土地利用类型为草地或沟道,沟道形态比较稳定且有少量的植被覆盖,有效减弱了侵蚀的发生。从平面曲率的影响上看,SHAP值随着平面曲率的增大逐渐增加,呈显著的正相关趋势。平面曲率越大,地面越凹陷,小于80时,地面较平坦,水流分散,冲刷力较弱,因此减弱了沟蚀;大于80时,地面较凸起,水流汇聚,流速加快,冲刷力增强,因此促进了沟蚀。从道路距离的影响上看,SHAP值整体分布接近0,但随着距离的增加,SHAP值略有下降趋势,尤其在300 m以

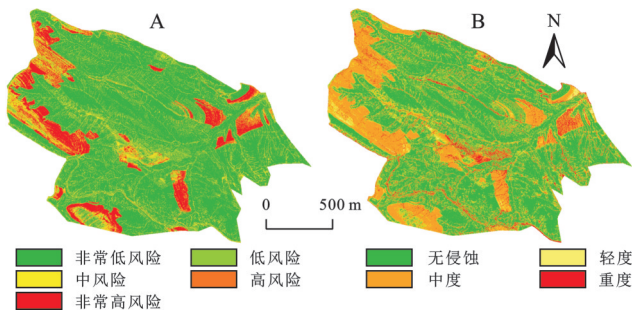


图9 50%样本密度下沟蚀风险和侵蚀等级图
Fig. 9 Gully erosion risk and erosion grade map under 50% sample density

综上所述,侵蚀强度不是简单地由整体风险等级决定,低风险区域仍可能发生中度甚至重度侵蚀,高风险区域也不一定全部为重度侵蚀。

4 结论

本研究利用随机森林模型,在4种样本密度(25%,50%,75%和100%)下,基于土地利用类型、坡度、平面曲率、坡向等9个特征因子对黄土高原典型小流域进行了沟蚀风险预测,研究得出以下结论:

内时影响较显著。靠近道路的区域风险相对较高,而远离道路的区域风险逐渐降低,这可能是因为道路附近土地扰动和人类活动较多,导致土壤侵蚀的风险增加。

3.4 沟蚀风险等级分析

表5为研究区的沟蚀风险和沟蚀严重程度的耦合分析表,揭示了区域内不同风险等级与侵蚀等级的分布特征和空间关系。研究区的沟蚀风险等级分为非常低风险、低风险、中风险、高风险和非常高风险5个等级,同时侵蚀沟严重程度分为无侵蚀、轻度侵蚀、中度侵蚀和重度侵蚀4个等级,如图9所示。研究结果表明,非常低风险区域占研究区面积的63.12%,主要分布在草地和稳定的沟道区域。低风险区域中,轻度侵蚀仅占0.85%,表明该区域中整体侵蚀强度较小。相比之下,中度、重度侵蚀分别占10.20%和3.90%,表明尽管为低风险区,但仍存在概率发生较为严重的沟蚀,此区域多处于沟道两侧,短时强降雨事件易在此区域汇流,特别是沟道两侧的低洼地带,由于坡度相对较小,径流容易聚集进而加剧沟蚀。对于中风险区域,中度侵蚀和重度侵蚀分别占4.81%和2.91%,此区域主要位于高程变化较大的地带,且以草地和耕地的过渡区为主,说明地形起伏加剧了侵蚀发生的风险。此外,高风险和非常高风险区域主要集中在耕地,尽管风险较高,侵蚀等级仍主要为中度侵蚀,作物根系增强了土壤稳定性,而作物覆盖减少了溅蚀的发生,从而在一定程度上降低了侵蚀强度。

表5 沟蚀风险和侵蚀严重程度耦合分析表

Table 5 Coupling analysis of gully erosion risk and erosion severity

侵蚀程度	%				
	非常低风险	低风险	中风险	高风险	非常高风险
无侵蚀	63.12	0	0	0	0
轻度侵蚀	0	0.85	0.34	0.39	0.32
中度侵蚀	0	10.20	4.81	4.71	6.97
重度侵蚀	0	3.90	2.91	0.92	0.56

(1) 在50%样本密度下,随机森林模型的准确率(0.901)、精确率(0.894)和Kappa C值(0.802)均为最高,同时能够准确识别各风险等级的空间分布,表明50%的样本密度是兼顾准确与效率的最优选择。

(2) 基于SHAP算法分析了50%样本密度下各特征因子对沟蚀风险预测的影响,结果表明土地利用类型是主导因子,平面曲率和坡度次之。

(3) 对沟蚀风险等级与侵蚀严重程度的分析表明,局部低风险区域也可能存在较高侵蚀等级(如中

度侵蚀),说明沟蚀风险等级不能代表侵蚀严重程度。

(4)暴雨条件下,黄土高原典型小流域的侵蚀沟主要分布在耕地区域,这一区域应作为沟蚀防治的重点。

参考文献 (References):

- [1] 赵景波,杜娟,黄春长.黄土高原侵蚀期研究[J].中国沙漠,2002,22(3):257-261.
Zhao J B, Du J, Huang C C. Study on erosion periods in Loess Plateau[J]. Journal of Desert Research, 2002, 22(3):257-261.
- [2] 张楠,张岩,王佳希,等.黄土丘陵沟壑区小流域侵蚀沟数量及形态特征[J].水土保持学报,2023,37(3):109-115.
Zhang N, Zhang Y, Wang J X, et al. Quantity and morphological parameters of gullies in small watersheds in the hilly-gully Loess Plateau [J]. Journal of Soil and Water Conservation, 2023, 37(3):109-115.
- [3] 王颢霖,焦菊英,唐柄哲,等.陕北子洲“7·26”暴雨后坡耕地细沟侵蚀及其影响因素分析[J].农业工程学报,2019,35(11):122-130.
Wang H L, Jiao J Y, Tang B Z, et al. Characteristics of rill erosion and its influencing factors in slope farmland after “7·26” rainstorm in Zizhou County, Shaanxi Province[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(11):122-130.
- [4] 杨力华,庞国伟,杨勤科,等.近50年来王茂沟流域侵蚀沟变化及其影响因素[J].水土保持学报,2020,34(2):64-70.
Yang L H, Pang G W, Yang Q K, et al. Changes and influencing factors of erosion gully in Wangmaogou watershed in the last 50 years[J]. Journal of Soil and Water Conservation, 2020, 34(2):64-70.
- [5] 万兆凯,宋庆臣,万普强,等.东北黑土区沟蚀研究进展与热点:基于CiteSpace计量分析[J].水土保持研究,2024,31(2):454-463.
Wan Z K, Song Q C, Wan P Q, et al. Research progress and hotspots of gully erosion in the black soil region of northeast: Bibliometric analysis based on CiteSpace [J]. Research of Soil and Water Conservation, 2024, 31(2):454-463.
- [6] Setargie T A, Tsunekawa A, Haregeweyn N, et al. Random Forest-based gully erosion susceptibility assessment across different agro-ecologies of the Upper Blue Nile basin, Ethiopia[J]. Geomorphology, 2023,431:108671.
- [7] Jiang C C, Fan W, Yu N Y, et al. Spatial modeling of gully head erosion on the Loess Plateau using a certainty factor and random forest model[J]. Science of the Total Environment, 2021,783:147040.
- [8] 曹文悦.基于机器学习的区域尺度沟蚀易发性评价:以元江干热河谷为例[D].昆明:云南大学,2023.
Cao W Y. Evaluation of regional scale gully erosion susceptibility based on machine learning: a case study of Yuanjiang dry-hot valley[D]. Kunming: Yunnan University, 2023.
- [9] Filho J P M, Guerra A J T, Cruz C B M, et al. Machine learning models for the spatial prediction of gully erosion susceptibility in the pirai drainage basin, paraiba do Sul middle valley, southeast Brazil[J]. Land, 2024,13(10):1665.
- [10] Huang D H, Su L, Zhou L L, et al. Assessment of gully erosion susceptibility using different DEM-derived topographic factors in the black soil region of Northeast China [J]. International Soil and Water Conservation Research, 2023,11(1):97-111.
- [11] Aboutaib F, Krimissa S, Pradhan B, et al. Evaluating the effectiveness and robustness of machine learning models with varied geo-environmental factors for determining vulnerability to water flow-induced gully erosion[J]. Frontiers in Environmental Science, 2023,11:1207027.
- [12] Amiri M, Pourghasemi H R, Ghanbarian G A, et al. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms [J]. Geoderma, 2019, 340:55-69.
- [13] 程军奇.近52年甘肃黄土高原极端降水事件的变化特征[D].兰州:西北师范大学,2013.
Cheng J Q. Variation characteristic of extreme precipitation in Loess Plateau of Gansu Province Region in recent 52 years[D]. Lanzhou: Northwest Normal University, 2013.
- [14] 黎珩,朱冰冰,边焯,等.1970—2020年黄土高原水蚀风蚀交错区极端降水时空变化研究及驱动因素分析[J].干旱区地理,2024,47(4):539-548.
Li H, Zhu B B, Bian H, et al. Temporal and spatial changes in extreme precipitation and its driving factors in the water-wind erosion crisscross region of the Loess Plateau from 1970 to 2020[J]. Arid Land Geography, 2024,47(4):539-548.
- [15] 杨丽娟,王春梅,张春妹,等.基于遥感影像研究极端暴雨条件下新成切沟发生发展规律[J].农业工程学报,2022,38(6):96-104.
Yang L J, Wang C M, Zhang C M, et al. Occurrence and development of newly formed gullies under extreme rainstorm conditions using remote sensing images [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022,38(6):96-104.
- [16] 陈卓鑫,王文龙,康宏亮,等.特大暴雨下不同土地利用类型坡面切沟发育特征[J].农业工程学报,2020,36(23):77-84.
Chen Z X, Wang W L, Kang H L, et al. Gully development characteristics of the slopes for different land-use types under extreme rainstorms [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36

- (23):77-84.
- [17] 刘焕永,杜鹏飞,赵莹,等.暴雨条件下黄土高原荞麦地的沟蚀特征:以陕西省定边县为例[J].水土保持研究,2024,31(5):1-8,17.
Liu H Y, Du P F, Zhao Y, et al. Characteristics of gully erosion on buckwheat field in the Loess Plateau under heavy rainfall conditions: A case study in Dingbian County, Shaanxi Province [J]. *Research of Soil and Water Conservation*, 2024,31(5):1-8,17.
- [18] 杨波,焦菊英,马晓武,等.2022年黄土高原典型暴雨侵蚀及洪水灾害调查分析[J].水土保持通报,2022,42(6):1-13.
Yang B, Jiao J Y, Ma X W, et al. Investigation and analysis of typical rainstorm erosion and flood disaster on Loess Plateau in 2022[J]. *Bulletin of Soil and Water Conservation*, 2022,42(6):1-13.
- [19] 刘宝元,杨扬,陆绍娟.几个常用土壤侵蚀术语辨析及其生产实践意义[J].中国水土保持科学,2018,16(1):9-16.
Liu B Y, Yang Y, Lu S J. Discriminations on common soil erosion terms and their implications for soil and water conservation[J]. *Science of Soil and Water Conservation*, 2018,16(1):9-16.
- [20] 王略,徐佳,董亚维.黄土高原典型小流域侵蚀沟形态和稳定性监测初探[J].水土保持通报,2018,38(1):126-130.
Wang L, Xu J, Dong Y W. Preliminary study on morphology and stability of erosion gully in typical small watersheds of Loess Plateau [J]. *Bulletin of Soil and Water Conservation*, 2018,38(1):126-130.
- [21] 王文娟,邓荣鑫,张树文.东北典型黑土区沟蚀发生风险评价研究[J].自然资源学报,2014,29(12):2058-2067.
Wang W J, Deng R X, Zhang S W. Preliminary research on risk evaluation of gully erosion in typical black soil area of Northeast China[J]. *Journal of Natural Resources*, 2014,29(12):2058-2067.
- [22] 李建军,陈玉兰,焦菊英,等.基于多元非线性空间建模的拉萨河流域沟蚀发生风险探测[J].农业工程学报,2022,38(17):73-82.
Li J J, Chen Y L, Jiao J Y, et al. Detecting gully occurrence risks using multivariate nonlinear spatial modeling in the Lhasa River Basin of China [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022,38(17):73-82.
- [23] 蒋根,肖诗荣,杨璇喆,等.基于机器学习的湖北省巴东县滑坡易发性评价[J].水利水电快报,2024,45(11):48-55.
Jiang G, Xiao S R, Yang X Z, et al. Vulnerability evaluation of landslide disaster in Badong County based on machine learning [J]. *Express Water Resources & Hydropower Information*, 2024,45(11):48-55.
- [24] 陈晨晨,武谦,张占友,等.澜沧江中下游流域土壤侵蚀时空演变特征[J].水土保持研究,2022,29(2):11-17,30.
Chen C C, Wu Q, Zhang Z Y, et al. Spatiotemporal change of soil erosion in the middle and lower reaches of Lancangjiang River [J]. *Research of Soil and Water Conservation*, 2022,29(2):11-17,30.
- [25] 王守华,王睿菘,孙希延,等.基于CF-CNN-LSTM模型的滑坡易发性评价[J].自然灾害学报,2024,33(5):84-95.
Wang S H, Wang R S, Sun X Y, et al. Landslide susceptibility evaluation based on CF-CNN-LSTM model [J]. *Journal of Natural Disasters*, 2024,33(5):84-95.
- [26] Chen W, Panahi M, Tsangaratos P, et al. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility [J]. *Catena*, 2019,172:212-231.
- [27] Kannangara K K P M, Zhou W H, Ding Z, et al. Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method [J]. *Journal of Rock Mechanics and Geotechnical Engineering*, 2022,14(4):1052-1063.
- [28] Wei Y J, Liu Z, Zhang Y, et al. Analysis of gully erosion susceptibility and spatial modelling using a GIS-based approach [J]. *Geoderma*, 2022,420:115869.

(上接第 100 页)

- [25] Li J W, Tian Y C, Wang D H, et al. Matching and driving mechanism analysis of the supply and demand relationships of soil conservation services in karst peak-cluster depression basin in Southwest Guangxi, China [J]. *Catena*, 2024,246:108438.
- [26] 陈田田,王钰茜,曾兴兰,等.西南地区生态系统服务关系特征及其与植被覆盖的约束效应[J].生态学报,2023,43(6):2253-2270.
Chen T T, Wang Y X, Zeng X L, et al. Characteristics and the constraint relationship between ecosystem services and vegetation coverage in the Southwest China [J]. *Acta Ecologica Sinica*, 2023,43(6):2253-2270.
- [27] Wu Q, Wang L, Wang T Y, et al. Spatial-temporal evolution analysis of multi-scenario land use and carbon storage based on PLUS-InVEST model: a case study in Dalian, China [J]. *Ecological Indicators*, 2024,166:112448.