第 58 卷 第 2 期 2025 年 (总 240 期)

北地 西 质

NORTHWESTERN GEOLOGY

Vol. 58 No. 2 2025(Sum240)



引文格式:李光明,杨玉飞,唐亚明,等.数据驱动模型评价滑坡易发性的对比研究:以黄河中游流域为例[J].西北 地质,2025,58(2):51-65. DOI: 10.12401/j.nwg.2024064

Citation: LI Guangming, YANG Yufei, TANG Yaming, et al. Comparison Study in Landslide Susceptibility Assessment by Using Data-driven models: A Case Study from the Middle Stream of the Yellow River[J]. Northwestern Geology, 2025, 58(2): 51–65. DOI: 10.12401/j.nwg.2024064

数据驱动模型评价滑坡易发性的对比研究: 以黄河中游流域为例

李光明1,杨玉飞2,3,唐亚明4,*,王小浩5,尹春旺4,冯凡4,周永恒4

(1. 天津市政工程设计研究总院有限公司,天津 300392;2. 浙江省地矿科技有限公司,浙江杭州 310007;
3. 河北工业大学 土木与交通工程学院,天津 300401;4. 中国地质调查局西安地质调查中心, 陕西 西安 710119;5. 陕西核工业工程勘察院有限公司,陕西 西安 710054)

摘 要:准确的滑坡易发性图有益于管理部门开展土地利用规划和防灾减灾工作,目前已经成为 了中国滑坡风险评估与管控的重点研究领域。本研究旨在对比分析不同数据驱动模型在区域滑 坡易发性评估中的表现,以黄河中游流域为研究区,通过详细的野外调查结合遥感图像视觉解 释,获得了包括684个历史滑坡点的数据库。选取了14个评价因子,利用Pearson相关系数分析 了这些因素之间的相关性,应用C5.0决策树算法确定了各因素的重要性。选取了3种典型的数 据驱动模型(加权信息量(WIV),支持向量机(SVM)和随机森林(RF))进行了区域滑坡易发性评 价,并通过受试者工作特征曲线(ROC)及其曲线下面积AUC值来验证模型的性能。结果表明, 距道路的距离、距河流的距离以及坡度是该地区滑坡发生最重要的贡献因素。大多数历史滑坡 都发生在滑坡易发性图中的中等和高易发区内。SVM和RF模型获得的高/极高易发区内的滑坡 点均超过总滑坡点的70%。RF模型表现最好,高易发性区占全区面积的21.9%,滑坡数量占全部 历史滑坡点的90.5%。AUC 精度的比较表明, RF模型的准确性高于其他两种模型: RF的AUC 为 0.904,而WIV和SVM的AUC 分别为0.845和0.847。

关键词: 滑坡易发性; 权重; 数据驱动模型; 决策树; 环境因子; 随机森林

中图分类号: P642.22 文献标志码: A 文章编号: 1009-6248(2025)02-0051-15

Comparison Study in Landslide Susceptibility Assessment by Using Data-driven models: A Case Study from the Middle Stream of the Yellow River

LI Guangming¹, YANG Yufei^{2,3}, TANG Yaming^{4,*}, WANG Xiaohao⁵, YIN Chunwang⁴, FENG Fan⁴, ZHOU Yongheng⁴

收稿日期: 2022-09-28; 修回日期: 2024-05-27; 责任编辑: 贾晓丹

基金项目: 国家重点研发计划课题"黄土高原基础设施密集区重大链生灾害信息共享技术平台研发及应用示范" (2023YFC3008405),陕西省科技创新团队项目"基于IT技术的地质灾害风险防控创新团队"(2023-CX-TD-33),天 津市规划和自然资源局科研项目"基于无人机遥感航测的天津北部蓟州山区地质灾害调查与稳定性综合评价及防 治措施研究"(津规科自筹 2022-40),"极端气象灾害条件下蓟州山区滑坡和泥石流风险评估与预警研究" (KJ[2024]25)联合资助。

作者简介:李光明(1984-),男,高级工程师,博士,主要研究方向为岩土工程。E-mail: liguangming20@126.com。

^{*}通讯作者: 唐亚明(1973-), 女, 正高级工程师, 博士, 主要研究方向为地质灾害风险评估。E-mail: tangyaming@mail.cgs.gov.cn。

(1. Tianjin Municipal Engineering Design & Research Institute, Tianjin 300392, China; 2. Zhejiang geology and mineral technology co. LTD,

Hangzhou 310007, Zhejiang, China; 3. School of Civil and Transportation Engineering, Hebei University of Technology,

Tianjin 300401, China; 4. Xi'an Center of China Geological Survey, Xi'an 710119, Shaanxi, China;

5. Shaanxi Nuclear Industry Engineering Survey Institute Co., Ltd. Xi'an 710054, Shaanxi, China)

Abstract: Accurate landslide susceptibility maps are beneficial for management departments to carry out land use planning and disaster prevention and mitigation. It has been an important field in the landslide risk assessment and management in China. This study aims to compare and analyze the performance of different data-driven models in the assessment of regional landslide susceptibility. The middle reaches of the Yellow river were selected as the study area, and a database including 684 historical landslide points was obtained through detailed field investigation combined with visual interpretation of remote sensing images. 14 evaluation factors were selected, Pearson correlation coefficient was used to analyze the correlation between these factors, and the C5.0 decision tree algorithm was used to determine the importance of each factor. Three typical data-driven models (Weighted Information Volume (WIV), Support Vector Machine (SVM) and Random Forest (RF)) were selected to evaluate the regional landslide susceptibility, and the performance of the models were verified by the Receiver Operating Characteristic (ROC) curve and the area AUC value under the curve. The results show that the distance from the road, the distance from the river and the slope are the most important contributing factors to the occurrence of landslides in this area. The majority of historical landslides occurred in the moderate and high susceptibility zones on the landslide susceptibility map. The landslide points in the high/very high susceptibility area obtained by SVM and RF models exceed 70% of the total landslide points. The RF model performed the best, with the high susceptibility area accounting for 21.9% of the area and the number of landslides accounting for 90.5% of all historical landslide points. A comparison of AUC accuracy shows that the RF model is more accurate than the other two models: RF has an AUC of 0.904, while WIV and SVM have AUCs of 0.845 and 0.847 respectively.

Keywords: landslide susceptibility; weight; data-driven model; decision tree; environmental factors; random forest

滑坡是中国最重要的自然灾害之一,每年都会造成一定的人员伤亡和经济损失(Froude et al., 2018; 郭子正等, 2018; 黄煜等, 2023)。滑坡易发性分析能够提供有关滑坡空间分布的有用信息,通常被选为滑坡风险评估的初始步骤。合理准确的滑坡易发性图能够帮助政府决策部门有效开展防灾减灾工作。因此,如何生成准确可靠的滑坡易发性图已成为现阶段的研究热点(郭子正等, 2019a; 张林梵, 2023; 林明明等, 2024)。

滑坡易发性评价方法大致可分为3类,包括专家 经验模型、基于物理的(或确定性)模型和数据驱动模 型(许冲等,2010; Goetz et al., 2011; Huang et al., 2017; Reichenbach et al., 2018; 郭子正等,2019b)。专家经验 模型通常属于定性方法,主要依赖于专家经验来描述 滑坡灾害(Sezer et al., 2017)。基于物理的模型将边坡 破坏机制和岩土参数相结合,通常以无限斜坡模型为 基础(He et al., 2021)。常见的基于物理的模型包括 TRIGRS(李婧等, 2022)、Scoops3D (He et al., 2021)、 SINMAP (武利, 2012)等。然而,在大范围内确定岩土 体准确的物理力学性质是一个挑战,通常具有很高的 不确定性。相比之下,数据驱动模型通过深入分析历 史滑坡与环境变量之间的统计或非线性关系来模拟 未来滑坡发生的可能性,通常比其他模型具有更高的 预测能力和客观性(Zèzere et al., 2017)。

受益于 GIS 和遥感技术的快速发展,数据驱动模型能够考虑各种类型的环境因子,包括地形地貌、水 文和地质因素等。然而,对于一个区域特定的滑坡数 据库,环境因子的最佳组合仍然是一个悬而未决的问题。一方面,研究者试图在分析中考虑更多因素,以 包含更多有用的信息;另一方面,必须在模型的复杂 度和性能中找到平衡,因为更多的因素虽然可能获得 较高的模型精度,但模型效率会随之降低。此外,很 少有研究讨论因素的地貌意义或解释单个因素在边 坡稳定性中的相关性(Segoni et al., 2020),且滑坡环境 因子的条件独立性也没有被广泛考虑(Pereira et al., 2012)。而考虑因素的相关性和独立性对于数据驱动 模型的质量非常重要。目前已经提出了一些方法来 研究滑坡易发性建模过程中的因子贡献程度,例如主 成分分析、信息增益比、前向消除和多重共线性技术 (张俊等, 2016; Binh Thai et al., 2019; Nsengiyumva et al., 2020; Tang et al., 2020)。笔者采用 Pearson 相关系 数计算因子间的独立性,并考虑采用 C5.0 决策树法评 价因子重要性。

前人研究比较了不同数据驱动模型的性能,包括 双变量、多变量和机器学习技术(Kouli et al., 2014; Bueechi et al., 2019;黄发明等, 2020;王本栋等, 2024)。 大多数比较结果表明,多变量技术优于双变量技术 (Rossi et al., 2010),而机器学习算法则优于两者(黄发 明等, 2021;李泽芝等, 2024;林琴等, 2024)。但它们 都缺乏因素独立性的考虑和重要性的比较。对于滑 坡多发地区,比较模型性能是必要的,因其有助于当 地的土地利用规划并减小滑坡风险。因此,笔者以黄 河流域中游为研究区,使用 3 种不同数据驱动方法 (WIV、SVM、RF)进行滑坡易发性评价并对比分析模 型性能,希望能够生成更加准确可靠的区域滑坡易发 性图。

1 研究区概况及数据准备

1.1 研究区概况

研究区(图 1)位于 E 110°30′0″~110°0′0″, N 37°10′0″~37°40′0″,橫跨陕西和山西两省。总面积 约为1661 km²,长约62 km,包括7个县。海拔从590 m 到1500 m 不等,中部海拔较低,因为黄河从中流过。 从地质上看,区内主要地质构造为王家汇背斜,方向 为 NW 30°。岩性单元包括寒武系至第四系的地层, 主要露头沉积物为黄土和砂岩和泥岩夹层。气候为 温带干旱气候,多年平均降水量为460 mm。降雨时 空变化明显:雨季一般从7月到9月,占年降雨量的 近70%,且南部地区的降雨量比北部地区多。

区内最具代表性的地貌单元是黄土高原,是中国 典型的滑坡多发地区之一(Zhang et al., 2010)。受当 地环境影响,植被覆盖范围小,坡体水土保持能力较 弱(Guo et al., 2020)。在这样的地质条件下,研究区近





年来发生了多起山体滑坡,造成了相当大的损失。但 目前还没有关于该地区利用数据驱动技术进行滑坡 易发性预测及制图的比较研究,因此是开展此类研究 的合适区域。

1.2 滑坡数据库

准确的滑坡数据库对区域易发性评估结果至 关重要,因为它能够提供滑坡的基本特征,例如位 置、面积、体积等(Fell et al., 2008)。此外,这是验证 结果的重要依据,因为研究者需要将实际的历史滑 坡点与预测结果进行比较(Huang et al., 2021)。本研 究通过遥感图像和野外实地调查相结合的方式确 定滑坡位置。卫星图像包括 LandsatTM8 和谷歌地 球图像。野外调查由中国地质调查局西安地质调 查中心于2018年进行。每一个滑坡位置坐标都由 便携式 GPS 记录下来,并重新投影到地图上,详细 信息则是根据滑坡报告确定的。然后将滑坡特征 存储在GIS属性表中,并与滑坡的空间位置相关联。 在最终的数据库中,共识别出滑坡684处,其中陕 西省414处,山西省270处。这些滑坡的深度(h)从 0.5 m 到 30 m 不等, 具体可分为浅层滑坡(h≤10 m)、 中深层滑坡(10 m<h≤25 m)和深层滑坡(h>25 m) 3个类型。其中,深层滑坡数量最少,仅占滑坡总数 的8%,中深层和浅层滑坡分别占35.8%和56.1%;体 积方面,约85%的滑坡体积小于1×10⁵m³,而体积大 于10⁶ m³的滑坡只有6个,均发育在黄河沿岸。依 据最新的滑坡分类方法(Hungr et al., 2014), 上述大

部分历史滑坡点都属于浅层土质滑坡,降雨和人类 活动是最常见的诱发原因。

1.3 环境因子的准备

选择输入因子是评估滑坡易发性的一项基本任务,因为它决定了建模过程中可以包含哪些信息 (Catani et al., 2013)。在本研究中,基于对研究区滑坡 机制的认知和前人相关文献,共考虑了14个因素作 为输入因子。这些因素可分为地形地貌因素、地质因 素、水文因素、环境因素和触发因素5类。

高程(图 2a)能反映滑坡势能,且不同高程区域的 人类工程活动存在差异,同时高程也会对边坡环境条 件(如气候、植被等)产生影响。研究区 DEM 数据来 源于开源网站地理空间数据云(http://www.gscloud. cn/),空间分辨率为30m。

坡度(图 2b)可以表示每个单元表面的陡峭程度, 对边坡稳定性有重要影响。它由 DEM 在 GIS 环境中 生成,范围为0到 59°。

坡向(图 2c)会导致小气候的差异(例如,阳光照 射和温度)并影响斜坡上的植被。使用 DEM 提取研 究区域的坡向,并划分为了 8 个方向(北、南、西、东、 西北、东北、西南、东南)和一个平坦区域(值为-1)。

平面曲率(图 2d)可以控制地表水流影响沉积和 侵蚀。该区域的平面曲率值从-2.825 到 4.227 不等。 剖面曲率(图 2e)控制流经斜坡的水流的加速和





(a).高程;(b).坡度;(c).坡向;(d).平面曲率;(e).剖面曲率;(f).地表粗糙度;(g).岩性;(h).地质构造;(i).NDWI;
 (j).到河流的距离;(k).NDVI;(l).土地利用;(m).降雨量;(n).到道路的距离

图2 研究区易发性评价的环境因子

Fig. 2 Environmental factors of landslide susceptibility assessment

减速,从而影响流动过程。它也是从 DEM 中计算得 到的,其结果为-3.908~4.199。

地表粗糙度(图 2f)是指地表面积与其投影面积 之比,是反映地表形态的宏观指标(Goetz et al., 2015)。 它是通过计算斜坡的表面积与垂直方向投影面积之 比而获得的。该因子的方程为:

$$SDS = 1/\cos(slope)$$
 (1)

岩性(图 2g)被认为是各种地质灾害的物质基础, 对控制滑坡的发展起着至关重要的作用(Catani et al., 2013)。研究区地质特征按地层年代和岩土性质划分 为砂壤土、黏土、红黏土、砂岩和石灰岩 5 个单元。

地质结构(图 2h)对于滑坡易发性评估至关重要, 是因为斜坡上的许多物理过程(如变形)与地质运动 有关;此外,地质构造也会影响斜坡上的节理和裂缝。

NDWI(图 2i)可以反映河流对地表的影响(Wang et al., 2020)。它主要来自 Landsat 8 TM 遥感图像,可以通过图像中不同波段的光谱反射率计算:

$$NDWI = \frac{P(Green) - P(NIR)}{P(Green) + P(NIR)}$$
(2)

式中: P(Green)代表绿色波段的光谱反射率; 而 P(NIR)是近红外波段的光谱反射率, 在本研究中, ND-WI值为-0.475~0.240。

与河流的距离(图 2j):河流可以切割和侵蚀河岸, 河流水位波动会重塑地形并影响斜坡的地下水。 与 到地质结构的距离类似,笔者计算了每个单元格到最 近河流的欧几里得距离,以表示河流的影响。

NDVI(图 2k)揭示了特定区域的植被覆盖程度,可以影响斜坡的水文过程。它也是从遥感图像中获得的。计算公式如下(Chang et al., 2020):

$$NDVI = \frac{P(NIR) - P(Red)}{P(NIR) + P(Red)}$$
(3)

式中: P(Red) 是红色波段的光谱反射率; P(NIR) 是近红外波段的光谱反射率。

土地利用(图 21)能够反映植被和人类工程活动的影响(Guo et al., 2025),还会涉及根系内聚力和水文

2025 年

过程。研究区主要有5种土地利用类型:水、居民区、 森林、草地和农田。

降雨(图 2m)是滑坡最常见的触发因素。在 GIS 中制作了研究区年平均降雨量图,显示南部地区的降 雨量高于北部地区。

到道路的距离(图 2n):研究区人口稠密,因此人 类活动也是引发滑坡的重要因素。该地区分布着多 条国道、高速公路和铁路,因此笔者以到公路的距离 作为反映人类活动强度的指标。

2 研究方法

2.1 因子独立性检验和重要性计算(C5.0决策树)

环境因子的独立性和重要性检验对易发性建模 有显著影响,能降低计算过程的复杂性,提高滑坡易 发性评价的准确性。如前文所述,文中环境因子的选 择主要基于专家经验和文献回顾(Catani et al., 2013; Guo et al., 2021)。上述 14 个环境因子虽然地质背景 不同,但与滑坡易发性存在内在关联。因此,首先采 用皮尔逊相关系数法计算所选因子的独立性。不同 因子对易发性制图的精度有不同的贡献,因此还需要 进行重要性计算来表明环境因子对滑坡的影响程度 (Zhu et al., 2022)。在本研究中,应用了 C5.0 决策树模 型(田乃满等, 2020)。C5.0 算法考虑了信息增益比 (*IGR*)方法来计算因素重要性,它允许基于概率度量 获得不确定性降低的水平。最佳解决方案是通过使 用最大 *IGR* 进行拆分计算的决策树向下增长来实现 的。*IGR*(GainRatio)可以表示如下:

$$GainRatio = \frac{Gains(S_q, m)}{Split(S_q, m)}$$
(4)

式中: *S_q* 是通过 Bootstrap 重采样随机选择的训练 数据集 D 的子集; *m* 是预测变量。 *Split*(*S_q*, *m*) 表示分 割信息。*Gains*(*S_q*, *m*)表示信息增益,可以表示为:

$$Gains(S_q, m) = H(S_q) - \sum_{v \in V(m)} \frac{|S_v|}{|S_q|} H(S_v)$$
(5)

式中: H(x)是数据集 x 的熵; V(m)是预测变量 m 的取值范围; S_V 是集合 S_a 的子集。

此外,提出了提升算法以提高模型的稳定性,并 减少分类错误(Dou et al., 2020)。采用交叉验证方法 来评估模型的性能,该方法有利于解决过拟合问题, 提高模型泛化能力(Yao et al., 2008)。使用了 R 软件 中内置包用于 C5.0 DT 建模,其中开发了一个函数,通 过计算落入训练数据集中的因素百分比来确定输入 数据的重要性。

2.2 滑坡易发性模型

2.2.1 加权信息量模型

信息量模型(IV)是一种常见的统计方法,它可以 整合专家主观经验和环境因子与滑坡相关性的客观 特征,已被广泛应用于滑坡易发性评估(Bhandary et al., 2013)。IV 模型首先计算各个评价因子不同等级 的信息量(Wang et al., 2019),表达式如下:

$$I(x_i, A) = \ln \frac{P(x_i/A)}{P(x_i)} \tag{6}$$

$$I(x_i, A) = \ln \frac{N_i/N}{S_i/S} = \ln \frac{N_i/S_i}{N/S}$$
(7)

式中: *S* 为区域总面积; *N* 为滑坡灾害数; *S_i* 是第 *i* 个因子的面积; *N_i* 为第 *i* 个因子中的滑坡灾害数。 *I_i*<0 表明该类别中滑坡发生的可能性小于稳定的可 能性,该因子对滑坡的贡献较小;反之 *I_i*>0 表明该类 别对于滑坡发生具有正向作用。然而,传统的信息量 模型没有考虑触发因素对滑坡的"贡献"差异。在 这项研究中,笔者应用层次分析法来改善这一缺陷, 并通过将每个因子的相关权重与其分类信息量的乘 积相加来计算加权信息值(WIV)。方程如下:

$$L = \sum_{i=1}^{n} W_i I_i \tag{8}$$

式中: W_i 为第i个因子的权重,L为滑坡易发性指数(LSI)。

2.2.2 随机森林模型

随机森林模型是近年来发展起来的一种算法,被 认为是组装了许多随机决策树。它的基本技术之一 是自举重采样技术,它在训练集中随机重复选择一些 样本来训练决策树,并生成其他树来组成随机森林。 在随机生成多棵决策树后,样本可以根据每棵决策树 的统计结果选择最佳分类(贾俊等,2023)。决策树的 形成需要叶子节点的完全分裂,每个叶子节点不能持 续分裂(Pradhan,2013)。大多数决策树算法都需要修 剪以避免过度拟合。在 RF 模型中随机选择样本可以 有效避免修剪分类结果(Goetz et al., 2015)。RF 模型 结合了每个决策树,可以代表最终的结果。它的优点 主要包括:可以在一定程度上避免过拟合,增强抗噪 能力;且无需特征值即可计算高维样本。

2.2.3 支持向量机模型

SVM 模型是一种用于统计数据二元分类的智能 算法(付泉等, 2024)。该模型通过核函数将环境因子 数据从原始空间映射到更高维空间。因此,样本在空间中是线性可分的,并能够分析滑坡点和非滑坡点的 正负分类之间的最大间距。在空间中搜索最优超平 面,分离样本,使样本间隔最大化,然后通过分类决策 函数对样本进行分类(Goetz et al., 2015)。{ x_i, y_i }是样 本数据的特征向量,其中 $i=1, 2, ..., l, x_i \in Rn, y_i \in \{-1, +1\}, 1$ 是样本个数,n表示输入维度。非线性映射 $\Phi(x)$ 将样本从输入空间映射到特征空间。分类超平面可 以计算为 $w\Phi(x)+b=0$,其中w, b是分类决策函数的 系数。最优超平面需要最大 2/IwI,可以转化为二次 规划问题。提出了拉格朗日乘子法来求解这个方程:

$$\begin{cases} \min \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i \\ s.t.y_i(w * x_i + b) \ge 1 - \xi_i \\ \xi_i \ge 0, i = 1, 2 \cdots n \end{cases}$$
(9)

式中: *ξ*_i 是松弛因子; *C* 是惩罚参数。简化的拉格朗日高维映射函数可以表示为:

$$K(x_{i}, x_{j}) = \phi(x_{i})\phi(x_{j})$$

$$\max Q(a) = \sum_{i=1}^{n} a_{i} - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i}a_{j}y_{i}y_{j}K(x_{i} \cdot x_{j})$$

$$s.t. \sum_{i=1}^{n} a_{i}y_{i} = 0, 0 \leq a_{i} \leq C$$
(10)

式中: *K*(*x_i*, *x_j*) 是核函数(Cherkassky, 1997)。则 SVR 模型可以建立为:

$$f(x) = sgn\left[\sum_{i=1}^{n} a_i * y_i K(x_i \cdot x) + b^*\right]$$
(11)

上述方法中提出了核函数,将高维空间问题简化 为低维空间问题。核函数的选择对于模型分类的准 确性具有重要意义。笔者选择 RBF(径向基函数)作 为核函数,因为它是应用最广泛的一种(Zhou et al., 2016)。

3 结果

3.1 易发性模型参数优化

文中区域滑坡易发性建模主要包括以下步骤:

(1)栅格文件的单元格分辨率设置为 30 m, 全区 共有 1 849 962 个栅格单元。在 GIS 环境下提出了格 点转折函数, 用于获取区域内所有环境因子的属性数 据。由于各个环境因子取值范围不同, 首先将各个因 素值归一化到 [0, 1] 区间。在建立环境因子体系后, 应用层次分析法计算各因子的权重。利用环境因子 判断矩阵计算了每个因素的归一化权重。不同统计 指标的结果(随机指数=1.58;一致性指数=0.092;一致 性比=0.058<0.1;λ_{max}=15.191)表明矩阵通过了一致性 检验。

(2)将 684 个滑坡点作为滑坡样本,在滑坡点周 围大于 500 m 的缓冲区内随机抽取非滑坡样本 684 个。 各个非滑坡点之间的距离大于 100 m。在 MATLAB 环境下,将滑坡和非滑坡样本随机划分为 80% 的训练 数据集和 20% 的测试数据集,利用不同的模型对训练 数据进行训练。然后将研究区所有栅格代入训练好 的模型中进行预测,得到了整个区域的易发性分布图。

(3)SVM 和 RF 模型在 Matlab 软件中实施以计算 滑坡易发性。具有 RBF 的 SVM 模型考虑了非负松弛 变量来确定最优超平面,改进了具有分类误差的样本 的优化,并模拟了训练集和测试集样本。k 折交叉验 证方法确定了两个关键参数,即惩罚因子C和核函数 参数 g, 最终得到最优参数 c = 0.7579 和 g = 0.584 3。 测试样本 1094 个, 预测成功 995 个, 失败 99 个, 训练 集样本准确率为 90.95%(图 3a)。测试样本 274 个,其 中预测成功 238个,失败 36个,测试样本准确率为 86.86%(图 3b)。RF 模型采用 bootstrap 的方式,从已 经放回的训练数据中抽取 500 个样本, 形成 500 棵决 策树。利用最优因子集作为模型运算决策树的生长 和分裂节点。将创建的 RF 分类器引入样本预测,结 果表明它具有良好的预测性能(图 4a)。测试样本 1092个,预测成功1090个,失败两个,训练集样本准 确率为 99.82%(图 4b)。测试样本 274个,其中预测 成功 246 个, 失败 28 个, 测试样本准确率为 87.97% (图 4c)。

3.2 环境因子重要性分析

采用 SPSS 软件中的双变量相关分析对 14 个环 境因子的独立性进行分析。相关系数 R[-1,1] 是计算 结果的判断指标。R 的绝对值越大,相关性越大。R= [-1,0),0,(0,1] 表示两个因素分别为负相关、无线性 相关、正相关,低相关(独立性)可通过|R| < 0.5 判断。 计算结果表明相关系数|R|max=0.486 < 0.5,说明环境 因子对滑坡灾害易发性评价影响不大,选择合理。

14个因子中的5个因子在滑坡易发性建模过程 中贡献较大,分别是到道路的距离(IM=1)、到河流的 距离(IM=0.97)、坡度(IM=0.93)、降雨量(IM=0.92), 海拔(IM=0.91)。岩性(IM=0.67)、土地利用(IM= 0.42)、地质构造(IM=0.32)和NDVI(IM=0.32)的贡献 中等(图5)。其他5个因素的贡献较低,包括NDWI、







图4 RF模型性能分析(a)、训练集预测结果(b)和测试集预测结果(c)

Fig. 4 (a) The RF model performance analysis, (b) training set prediction results and (c) test set prediction results

平面曲率、轮廓曲率、表面粗糙度和坡向。重要性计 算结果与层次分析法计算的权重一致,其中与道路的 距离(W=0.159)、与河流的距离(W=0.136)、降雨量 (W=0.132)、坡度(W=0.132)具有较高的权重,验证了 主观权重结果的合理性和模型计算的客观重要性。 总体而言,结果表明,本研究区域滑坡发生的两个触 发因素(到河流的距离和降雨量)比其他因素更重要。 此外,没有任何因子具有负重要值,因此所有因子都可以合理地纳入滑坡易发性评价。

依据公式(4)、(5)计算 GIS 环境下各环境因子不同等级的信息量值,再乘以因素权重得到各因素的WIV 和排名。表1总结了在特定环境因子内分布的滑坡数量与总数的比值(Ni/N),在特定环境因子内分布的滑坡面积与整个区域的比值(Si/S), IV(I)、权重、







WIV,以及每个因子的每个级别的排名。根据结果可 知,该区域的 IV 值为-2.11~2.47。排名前 6 的因子 的 WIV 分别为: 道路距离(0~100 m)、降雨量(> 475 mm)、河流距离(0~100 m)、坡度(23°~59°)、 NDWI(0.008~0.240), 土地使用(定居)。这些因素显 着影响研究地点历史滑坡的发展。

3.3 不同滑坡易发性图中的滑坡分布

表1中计算的各因子不同范围的 WIV 由 ArcGIS 赋值,并由 ArcGIS 栅格计算器叠加,以确定该地区 的滑坡易发性指数(LSI)。LSI值为-0.838~0.940 (图 6a)。将 SVM 和 RF 方法预测的 LSI 通过 ArcGIS 软件分配到研究区对应的格网中,得到对应的滑坡 易发性图。来自 SVM 和 RF 模型的研究区域 LSI 值

滑坡环境因子 IV 的计算结果 表 1

Tab. 1Calculation of the IV of the landslide environmental factors												
环境因子	值	Ni/N	Si/S	密度比	信息量	权重	加权信息量	排名				
高程 (m)	594~774	0.39	0.39	1.00	0.90		0.065 7	9				
	$774 \sim 901$	0.22	0.76	0.29	-0.33		0.024 09	61				
	$901\!\sim\!1028$	0.26	0.69	0.38	-0.07	0.073	0.005 11	40				
	$1028\!\sim\!1183$	0.11	0.51	0.21	-0.67		0.048 91	66				
	$1183\!\sim\!1510$	0.01	0.08	0.16	-0.92		0.067 16	68				
坡度 (°)	$0\!\sim\!7$	0.14	0.19	0.74	-0.30		-0.032 1	62				
	$7\sim 12$	0.20	0.29	0.68	-0.39		0.041 73	64				
	$12 \sim 17$	0.26	0.28	0.96	-0.05	0.107	0.005 35	42				
	17~23	0.25	0.18	1.39	0.33		0.035 31	15				
	23~59	0.14	0.06	2.44	0.89		0.095 23	4				
	北(0~22.5)	0.09	0.06	1.56	0.45		0.013 95	22				
坡向 (°)	东北(22.5~67.5)	0.13	0.12	1.13	0.12		0.003 72	28				
	东(67.5~112.5)	0.15	0.13	1.14	0.13		0.004 03	27				
	东南(112.5~157.5)	0.10	0.12	0.79	-0.23		0.007 13	44				
	南(157.5~202.5)	0.12	0.13	0.94	-0.06	0.031	0.001 86	34				
	西南(202.5~247.5)	0.10	0.15	0.68	-0.38		0.011 78	51				
	西(247.5~292.5)	0.09	0.14	0.68	-0.39		0.012 09	52				
	西北(292.5~337.5)	0.14	0.11	1.27	0.24		0.007 44	25				
	北(337.5-360)	0.08	0.05	1.57	0.45		0.013 95	22				
平面曲率	$-2.824 \sim -0.345$	0.06	0.06	1.00	-0.16		0.004 96	38				
	$-0.345 \sim -0.097$	0.22	0.25	0.88	-0.12		0.003 72	36				
	$-0.097 \sim 0.095$	0.35	0.34	1.03	0.03	0.031	0.000 93	31				
	$0.095 \sim 0.343$	0.31	0.27	1.13	0.12		0.003 72	28				
	$0.343 \sim 4.227$	0.07	0.07	1.00	-0.10		-0.003 1	35				
剖面曲率	$-3.908 \sim -0.393$	0.05	0.06	0.85	-0.16		0.004 96	38				
	$-0.393 \sim -0.140$	0.16	0.20	0.79	-0.23		0.007 13	44				
	$-0.140 \sim 0.082$	0.28	0.37	0.77	-0.26	0.031	0.008 06	47				
	$0.082 \sim 0.367$	0.36	0.29	1.23	0.20		0.006 2	26				
	$0.367 \sim 4.199$	0.15	0.08	1.84	0.61		0.018 91	18				

								续表1
	值	Ni/N	Si/S	密度比	信息量	权重	加权信息量	排名
	1~1.023	0.34	0.48	0.69	-0.37		0.015 54	55
	$1.023 \sim 1.052$	0.33	0.32	1.02	0.02		0.000 84	32
地表粗糙度	$1.052 \sim 1.097$	0.21	0.15	1.41	0.35	0.042	0.014 7	21
	1.097~1.199	0.09	0.04	2.23	0.80		0.033 6	16
	1.199~1.919	0.03	0.01	3.00	2.15		0.090 3	8
	沙壤土	0.16	0.19	0.88	-0.13		0.006 76	43
	黏土	0.33	0.43	0.77	-0.26		0.013 52	53
岩性	红黏土	0.06	0.03	2.00	0.50	0.052	0.026	17
	砂岩	0.43	0.31	1.38	0.32		0.016 64	19
	石灰岩	0.02	0.04	0.49	-0.71		0.036 92	63
	$0 \sim 2\ 709.069$	0.37	0.28	1.30	0.26		0.012 74	24
	$2709.069 \sim 5727.746$	0.29	0.28	1.03	0.03		0.001 47	30
距断层距离 (m)	5 727.746~ 9 056.030	0.16	0.19	0.83	-0.18	0.049	0.008 82	48
	$9056.030{\sim}13003.531$	0.11	0.16	0.72	-0.32		0.015 68	56
	$13003.531{\sim}19814.904$	0.07	0.09	0.79	-0.23		0.011 27	50
	$-0.475 \sim -0.235$	0.06	0.10	0.56	-0.58		0.022 04	60
	$-0.235 \sim -0.196$	0.25	0.37	0.69	-0.38		0.014 44	54
NDWI	$-0.196 \sim -0.151$	0.47	0.43	1.09	-0.02	0.038	0.000 76	33
	$-0.151 \sim 0.008$	0.13	0.06	2.19	1.07		0.040 66	11
	$0.008\!\sim 0.240$	0.09	0.04	2.27	2.47		0.093 86	5
	$-0.198 \sim 0.008$	0.02	0.01	2.00	1.13		0.042 94	10
	$0.008 \sim 0.135$	0.31	0.20	1.53	0.43		0.016 34	20
NDVI	$0.135 \sim 0.180$	0.35	0.40	0.88	-0.13	0.038	0.004 94	37
	0.180~0.235	0.26	0.30	0.87	-0.14		0.005 32	41
	0.235~0.536	0.05	0.09	0.57	-0.57		0.021 66	59
	$0\!\sim\!100$	0.35	0.15	2.33	0.85		0.115 6	3
	$100\sim 200$	0.16	0.12	1.32	0.28		0.038 08	14
距河流距离(m)	$200 \sim 300$	0.11	0.13	0.86	-0.15	0.136	-0.0204	58
	$300\sim500$	0.14	0.21	0.66	-0.41		0.055 76	67
	$500 \sim 1776.851$	0.23	0.38	0.60	-0.51		0.069 36	69
	$0\!\sim\!100$	0.72	0.16	4.56	1.52		0.241 68	1
	$100\sim 200$	0.12	0.12	0.95	-0.05		0.007 95	46
距道路距离(m)	$200\sim 300$	0.05	0.12	0.41	-0.90	0.159	-0.143 1	71
	$300\sim500$	0.06	0.18	0.34	-1.07		0.170 13	72
	$500 \sim 2835.437$	0.05	0.42	0.12	-2.11		0.335 49	74
	水面	0.03	0.02	1.65	0.50		0.04	12
	村庄	0.17	0.05	3.22	1.17		0.093 6	6
土地利用	林地	0.24	0.28	0.87	-0.14	0.08	-0.011 2	49
	草地	0.28	0.48	0.58	-0.54		-0.043 2	65
	农田	0.29	0.17	1.66	0.50		0.04	12
	< 400	0.10	0.22	0.47	-1.45		-0.1914	73
	$400 \sim 425$	0.22	0.11	1.99	0.69		0.091 08	7
降雨 (mm)	425~450	0.12	0.13	0.87	-0.14	0.132	0.018 48	57
	450~475	0.10	0.22	0.44	-0.82		0.108 24	70
	> 475	0.46	0.33	1.40	1.57		0.207 24	2

分别为-4.236~1.807(图 6b)和0~1(图 6c)。采用自 然断点法将LSI 划分为高、中、低、极低4个易发性 等级,得到了3种数据驱动模型的易发性分布图 (图 7)。结果表明,3个模型得到的高易发地带与历 史滑坡点的空间分布基本一致。Ni/N和Si/S值都随 着易发性水平的增加而增加,从而表明这些模型对 滑坡易发区/非滑坡区的预测较好(图 8)。WIV 模型 的每一层在总面积中所占的比例都比较均匀。高易 发性面积占总面积的10.8%,滑坡点数占总点数的 57.8%。相比之下,易发性极低的地区占整个区域 的33.2%,但滑坡数量仅占所有点的2.5%。对于高 易发性区域识别的滑坡数量,SVM和RF模型均超 过70%,表明大部分滑坡预测成功。RF表现最好: 高易发性区占全区面积的21.9%,滑坡个数占90.5%, 极低易发性区面积占总面积的39.4%,滑坡个数占 总面积只有 0.7%。这些结果表明, RF 方法更好地预测了高易发性地区的历史滑坡, 生成的滑坡易发性 图更有效。

3.4 模型验证与比较

模型验证对于模型适应度评估和生成的滑坡易 发性图的质量评估都至关重要(Guzzetti et al., 2005)。 笔者提出了基于混淆矩阵的统计指标来评估这些模 型的预测能力,并生成了接收者操作特征(ROC)。笔 者使用了接收者操作特征曲线(ROC)来定量评估及 对比这些模型的性能,3个模型的 ROC 曲线下面积 (AUC)均在 0.8 以上,说明这些模型在滑坡易发性评 估中均具有良好的预测能力(图 9)。RF 模型的 AUC 值为 0.903 5, 而 SVM 和 WIV 的 AUC 值分别为 0.847 2、 0.844 6。SVM 和 RF 模型得到的极低易发性区内的滑 坡点只有 5 个,占清单点的 0.7%。通过上述对比可知,







图7 从不同模型获得的滑坡易发性图: WIV 模型(a), SVM 模型(b)和 RF 模型(c) Fig. 7 Landslide susceptibility maps obtained from different models: (a) WIV model, (b) SVM model and (c) RF model





中

Fig. 8 Statistical indicators of the historical landslides in each susceptibility level: (a) WIV model, (b) SVM model and (c) RF model

低



0

极低

图9 滑坡易发性评估中三种数据驱动模型的 ROC 曲线及精度



SVM 和 RF 模型的验证结果更加客观准确。虽然 WIV 模型可以量化因素与滑坡发生之间的关系,但专 家判断主要是非定性和经验性的。因此,误差高于机

器学习模型。在这3种模型中,随机森林的预测能力 最高。

0

高

2025年

结论 4

(1)分别使用了3种数据驱动模型生成黄河中游 流域的区域滑坡易发性图。环境因子间的皮尔逊相 关系数表明,所有选择的14个因素都是条件独立的。 从 C5.0 决策树方法获得的结果揭示了该地区滑坡发 生最关键的因素,即到道路的距离、到河流的距离和 坡度。

(2)使用 ROC 曲线的验证和比较结果表明, RF 模 型(AUC=0.904)比 WIV(AUC=0.845)和 SVM(AUC= 0.847)模型具有更高的准确性。RF模型确定了97.1% 的历史滑坡清单点位于高易发性区域,而只有0.7% 的滑坡位于低易发性区域。

(3)3种技术已被确认为中国黄土高原滑坡易发 性评估较有前景的模型,特别是 RF 模型。研究结果 第2期

有助于未来在具有相似地质和环境的地区进行滑坡 风险管理和防灾减灾工作。

参考文献(References):

- 付泉,党光普,李致博,等.基于分形维数耦合支持向量机和熵 权模型的滑坡易发性研究[J].西北地质,2024,57(6): 255-267.
- FU Quan, DANG Guangpu, LI Zhibo, et al. Study of Landslide Susceptibility Mapping Based on Fractal Dimension Integrating Support Vector Machine with Index of Entropy Model[J]. Northwestern Geology, 2024, 57(6): 255–267.
- 郭子正,殷坤龙,黄发明,等.基于地表监测数据和非线性时间 序列组合模型的滑坡位移预测[J].岩石力学与工程学报, 2018,37(S1):3392-3399.
- GUO Zizheng, YIN Kunlong, HUANG Faming, et al. Landslide displacement prediction based on surface monitoring data and nonlinear time series combination model[J]. Chinese Journal of Rock Mechanics and Engineering, 2018, 37(S1): 3392–3399.
- 郭子正,殷坤龙,黄发明,等.基于滑坡分类和加权频率比模型 的滑坡易发性评价[J].岩石力学与工程学报,2019a,38(2): 287-300.
- GUO Zizheng, YIN Kunlong, HUANG Faming, et al. Evaluation of landslide susceptibility based on landslide classification and weighted frequency ratio model[J]. Chinese Journal of Rock Mechanics and Engineering, 2019a, 38(2): 287–300.
- 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, 2019b, 44(12): 4299-4312.
- GUO Zizheng, YIN Kunlong, FU Sheng, et al. Evaluation of landslides susceptibility based on GIS and WOE-BP model[J]. Earth Science, 2019b, 44(12): 4299–4312.
- 黄发明, 叶舟, 姚池, 等. 滑坡易发性预测不确定性: 环境因子不同属性区间划分和不同数据驱动模型的影响[J]. 地球科学, 2020, 45(12): 4535-4549.
- HUANG Faming, YE Zhou, YAO Chi, et al. Uncertainties of Landslide Susceptibility Prediction: Different Attribute Interval Divisions of Environmental Factors and Different Data-Based Models[J]. Earth Science, 2020, 45(12): 4535–4549.
- 黄发明,陈佳武,唐志鹏,等.不同空间分辨率和训练测试集比 例下的滑坡易发性预测不确定性[J].岩石力学与工程学 报,2021,40(6):1155-1169.
- HUANG Faming, CHEN Jiawu, TANG Zhipeng, et al. Uncertainties of Landslide Susceptibility Prediction: Different Attribute Interval Divisions of Environmental Factors and Different Data-Based Models[J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(6): 1155–1169.

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.

- HUANG Yu, XIE Wanli, LIU Qiqi, et al. Landslide Susceptibility Assessment Based on GIS and MaxEnt Model: Example from Central Districts in Tongchuan City[J]. Northwestern Geology, 2023, 56(1): 266–275.
- 贾俊, 毛伊敏, 孟晓捷, 等. 深度随机森林和随机森林算法的滑 坡易发性评价对比—以汉中市略阳县为例[J]. 西北地质, 2023, 56(3): 239-249.
- JIA Jun, MAO Yimin, MENG Xiaojie, et al. Comparison of Landslide Susceptibility Evaluation by Deep Random Forest and Random Forest Model: A Case Study of Lueyang County, Hanzhong City[J]. Northwestern Geology, 2023, 56(3): 239–249.
- 李婧, 卢玲, 唐泽. 基于 TRIGRS 模型的区域降雨型浅层滑坡危险性评价[J]. 甘肃水利水电技术, 2022, 58: 24-27.
- LI Jing, LU Ling, TANG Ze. Risk assessment of regional rainfalltype shallow landslide based on TRIGRS model[J]. Gansu Water Resources and Hydropower Technology, 2022, 58: 24–27.
- 李泽芝, 王新刚. 镇域尺度下秦巴山区堆积层滑坡易发性不同 单元评价性能对比研究[J]. 西北地质, 2024, 57(1): 1-11.
- LI Zezhi, WANG Xingang. Comparative Study on Evaluation Performance of Different Units of Susceptibility of Accumulation Layer Landslide in Qinba Mountain Area at Town Scale[J]. Northwestern Geology, 2024, 57(1): 1–11.
- 林琴,郭永刚,吴升杰,等.基于梯度提升的优化集成机器学习 算法对滑坡易发性评价:以雅鲁藏布江与尼洋河两岸为例 [J].西北地质,2024,57(1):12-22.
- LIN Qin, GUO Yonggang, WU Shengjie, et al. Evaluation of Landslide Susceptibility by Optimization Integrated Machine Learning Algorithm Based on Gradient Boosting: Take Both Banks of Yarlung Zangbo River and Niyang River as Examples[J]. Northwestern Geology, 2024, 57(1): 12–22.
- 林明明, 赵勇, 王坤, 等. 基于多源时序 InSAR 技术的滑坡隐患 早期识别[J]. 西北地质, 2024, 57(6): 268-277.
- LIN Mingming, ZHAO Yong, WANG Kun, et al. Early Identification of Potential Dangers of Loess Landslide Based on Multi-Source and Time Series InSAR[J]. Northwestern Geology, 2024, 57(6): 268–277.
- 田乃满, 兰恒星, 伍宇明, 等. 人工神经网络和决策树模型在滑坡易发性分析中的性能对比[J]. 地球信息科学学报, 2020, 22(12): 2304-2316.
- TIAN Naiman, LAN Hengxing, WU Yuming, et al. Performance Comparison of BP Artificial Neural Network and CART Decision Tree Model in Landslide Susceptibility Prediction[J]. Journal of Geo-Information Science, 2020, 22(12): 2304–2316.
- 王本栋,李四全,许万忠,等.基于3种不同机器学习算法的滑 坡易发性评价对比研究[J].西北地质,2024,57(1):34-43.
- WANG Bendong, LI Siquan, XU Wanzhong, et al. A Comparative

Study of Landslide Susceptibility Evaluation Based on Three Different Machine Learning Algorithms [J]. Northwestern Geology, 2024, 57(1); 34–43.

- 武利. 基于 SINMAP 模型的区域滑坡危险性定量评估及模型验证[J]. 地理与地理信息科学, 2012, 28(2): 35-39+113.
- WU Li. Quantitative assessment and model validation of regional landslide risk based on SINMAP model[J]. Geography and Geographic Information Science, 2012, 28(2): 35–39+113.
- 许冲,戴福初,姚鑫,等.基于 GIS 与确定性系数分析方法的汶 川地震滑坡易发性评价[J].工程地质学报,2010,18(1): 15-26.
- XU Chong, DAI Fuchu, YAO Xin, et al. Landslide susceptibility evaluation of Wenchuan earthquake based on GIS and deterministic coefficient analysis method[J]. Chinese Journal of Engineering Geology, 2010, 18(1): 15–26.
- 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.
- ZHANG Jun, YIN Kunlong, WAND Jiajia, et al. Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir[J]. Chinese Journal of Rock Mechanics and Engineering, 2016, 35(2): 284–296.
- 张林梵. 基于时序 InSAR 的黄土滑坡隐患早期识别—以白鹿塬 西南区为例[J]. 西北地质, 2023, 56(3): 250-257.
- ZHANG Linfan. Early Identification of Hidden Dangers of Loess Landslide Based on Time Series InSAR: A Case Study of Southwest Bailuyuan[J]. Northwestern Geology, 2023, 56(3): 250-257.
- Bhandary N P, Dahal R K, Timilsina M, et al. Rainfall event-based landslide susceptibility zonation mapping[J]. Natural Hazards, 2013, 69(1): 365–388.
- Binh T P, Jaafari A, Prakash I, et al. A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling[J]. Bulletin of Engineering Geology and the Environment, 2019, 78(4): 2865–2886.
- Bueechi E, Klimes J, Frey H, et al. Regional-scale landslide susceptibility modelling in the Cordillera Blanca, Perua comparison of different approaches[J]. Landslides, 2019, 16(2): 395–407.
- Catani F, Lagomarsino D, Segoni S, et al. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues[J]. Natural Hazards and Earth System Sciences, 2013, 13(11): 2815–2831.
- Chang Z, Du Z, Zhang F, et al. Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models[J]. Remote Sensing, 2020, 12(3): 502.
- Cherkassky V. The nature of statistical learning theory[J]. IEEE transactions on neural networks, 1997, 8(6): 1564.
- Dou J, Yunus A P, Dieu T B, et al. Improved landslide assessment

using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan[J]. Landslides, 2020, 17(3): 641–658.

- Fell R, Cororninas J, Bonnard C, et al. Guidelines for landslide susceptibility, hazard and risk-zoning for land use planning[J]. Engineering Geology, 2008, 102(3-4): 85–98.
- Froude M J, Petley D N. Global fatal landslide occurrence from 2004 to 2016[J]. Natural Hazards and Earth System Sciences, 2018, 18(8): 2161–2181.
- Goetz J N, Brenning A, Petschko H, et al. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling[J]. Computers & Geosciences, 2015, 81: 1–11.
- Goetz J N, Guthrie R H, Brenning A. Integrating physical and empirical landslide susceptibility models using generalized additive models[J]. Geomorphology, 2011, 129(3-4): 376–386.
- Guo Z, Shi Y, Huang F, et al. Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management[J]. Geoscience Frontiers, 2021, 12(6): 101249.
- Guo Z, Yin K, Liu Q, et al. Rainfall Warning of Creeping Landslide in Yunyang County of Three Gorges Reservoir Region Based on Displacement Ratio Model[J]. Earth Science, 2020, 45(2): 672–684.
- Guo Z, Wang H, He J, et al. PSLSA v2.0: An automatic Python package integrating machine learning models for regional landslide susceptibility assessment[J]. Environmental Modelling & Software, 2025, 186: 106367.
- Guzzetti F, Reichenbach P, Cardinali M, et al. Probabilistic landslide hazard assessment at the basin scale[J]. Geomorphology, 2005, 72(1-4): 272–299.
- He J, Qiu H, Qu F, et al. Prediction of spatiotemporal stability and rainfall threshold of shallow landslides using the TRIGRS and Scoops3D models[J]. Catena, 2021, 197: 104999.
- Huang F, Yin K, Huang J, et al. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine[J]. Engineering Geology, 2017, 223: 11–22.
- Huang Y, Xu C, Zhang X, et al. An Updated Database and Spatial Distribution of Landslides Triggered by the Milin, Tibet M(w)6.4 Earthquake of 18 November 2017[J]. Journal of Earth Science, 2021, 32(5): 1069–1078.
- Hungr O, Leroueil S, Picarelli L. The Varnes classification of landslide types, an update[J]. Landslides, 2014, 11(2): 167–194.
- Kouli M, Loupasakis C, Soupios P, et al. Landslide susceptibility mapping by comparing the WLC and WofE multi-criteria methods in the West Crete Island, Greece[J]. Environmental Earth Sciences, 2014, 72(12): 5197–5219.
- Nsengiyumva J B, Valentino R. Predicting landslide susceptibility and risks using GIS-based machine learning simulations, case of

upper Nyabarongo catchment[J]. Geomatics Natural Hazards & Risk, 2020, 11(1): 1250–1277.

- Pereira S, Zezere J L, Bateira C. Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models[J]. Natural Hazards and Earth System Sciences, 2012, 12(4): 979–988.
- Pradhan B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS[J]. Computers & Geosciences, 2013, 51: 350–365.
- Reichenbach P, Rossi M, Malamud B D, et al. A review of statistically-based landslide susceptibility models[J]. Earth-Science Reviews, 2018, 180: 60–91.
- Rossi M, Guzzetti F, Reichenbach P, et al. Optimal landslide susceptibility zonation based on multiple forecasts[J]. Geomorphology, 2010, 114(3): 129–142.
- Segoni S, Pappafico G, Luti T, et al. Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization[J]. Landslides, 2020, 17(10): 2443–2453.
- Sezer E A, Nefeslioglu H A, Osna T. An expert-based landslide susceptibility mapping (LSM) module developed for Netcad Architect Software [J]. Computers & Geosciences, 2017, 98: 26–37.
- Tang Y, Feng F, Guo Z, et al. Integrating principal component analysis with statistically-based models for analysis of causal factors and landslide susceptibility mapping: A comparative

study from the loess plateau area in Shanxi (China)[J]. Journal of Cleaner Production, 2020, 277: 124159.

- Wang Q, Guo Y, Li W, et al. Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor[J]. Geomatics Natural Hazards & Risk, 2019, 10(1): 820–835.
- Wang Y, Feng L, Li S, et al. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China[J]. Catena, 2020, 188: 104425.
- Yao X, Tham L G, Dai F. 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.
- Zêzere J L, Pereira S, Melo R, et al. Mapping landslide susceptibility using data-driven methods[J]. Science of The Total Environment, 2017, 589: 250–267.
- Zhang M, Liu J. Controlling factors of loess landslides in western China[J]. Environmental Earth Sciences, 2010, 59(8): 1671– 1680.
- Zhou C, Yin K, Cao Y, et al. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China[J]. Engineering Geology, 2016, 204: 108–120.
- Zhu L, Wang G, Huang F, et al. Landslide Susceptibility Prediction Using Sparse Feature Extraction and Machine Learning Models Based on GIS and Remote Sensing [J]. Ieee Geoscience and Remote Sensing Letters, 2022, 19: 1–5.