

基于人工智能的滑坡灾害区域预警模型研究

周晴莎, 李同翔, 于全伟

中石化天然气分公司四川天然气销售中心, 四川 成都

收稿日期: 2024年11月19日; 录用日期: 2025年1月17日; 发布日期: 2025年1月26日

摘要

我国受地质灾害危害严重, 采取区域预警是降低灾害带来损失的重要手段之一, 其中预警模型是预警工作的关键。本文采用人工智能方法, 基于历史灾害数据、地质环境数据及诱发因素数据, 通过训练样本集构建、模型训练与优化评估等环节, 构建基于机器学习的滑坡灾害区域预警模型, 模型准确率为0.934, AUC值为0.95, 模型准确率及泛化能力均较好。通过实例校验, 人工智能预警模型相较统计预警模型预警结果准确率有明显提高, 预警结果的刻画也要更加精细, 对滑坡灾害区域风险精细化防控有较好的指导作用。

关键词

滑坡灾害, 区域预警, 人工智能, 精细化防控

Research on Regional Early Warning Model of Landslide Disaster Based on Artificial Intelligence

Qingsha Zhou, Tongxiang Li, Quanwei Yu

Sichuan Natural Gas Sales Center of SINOPEC Gas Company, Chengdu Sichuan

Received: Nov. 19th, 2024; accepted: Jan. 17th, 2025; published: Jan. 26th, 2025

Abstract

China is seriously affected by geological disasters. Taking regional early warning is one of the important means to reduce the losses caused by disasters, and the early warning model is the key to early warning work. This paper adopts artificial intelligence methods, based on historical disaster data, geological environment data and inducing factor data, through training sample set construction, model training and optimization evaluation, to construct a regional early warning model for

landslide disasters based on machine learning. The model accuracy is 0.934, the AUC value is 0.95, and the model accuracy and generalization ability are good. Through example verification, the accuracy of the artificial intelligence early-warning model is obviously improved compared with the statistical early-warning model, and the description of the early-warning results should be more precise, it has a good guiding role in the refined prevention and control of regional risks of landslide disasters.

Keywords

Landslide Disaster, Regional Early Warning, Artificial Intelligence, Refined Prevention and Control

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中国是世界上地质灾害分布最广、危害最严重的国家之一。地质灾害平均每年造成数百人死亡、数十亿元直接经济损失。我国自 2003 年起开始分级开展区域地质灾害气象预警业务[1], 取得了较好的防灾减灾成效, 其中预警模型是该业务的关键。

目前, 国内区域地质灾害预警模型主要采用统计预警模型, 具体又可细分为隐式统计模型、显式统计模型和动力预报模型[2], 这些模型在各地开展预警业务都取得了一定成效[3][4]。近年来, 随着人工智能技术的蓬勃发展, 基于人工智能的机器学习和深度学习算法日益成熟, 逐渐有学者将人工智能应用于区域地质灾害预警模型。刘艳辉等人以四川省青川县为研究对象, 在 2018 年和 2020 年分别采用统计方法和机器学习方法构建预警模型, 选取了 2018 年 6 月 26 日青川县日常预警业务实例作为校验, 机器学习模型相对统计模型预警效果更好[5][6]。李阳春、刘艳辉等人分别采用 BP 神经网络和随机森林在贵州和福建地区建立区域预警模型, 模型都取得了较好的成效[7][8]。谢旭阳、黄桥等人分别针对矿区和管道开展的人工智能区域预警模型研究, 都为工程安全保障提供了科学指导[9][10]。孙德亮采用机器学习方法分析了这些因素与滑坡发生的关系, 并构建了滑坡易发性区划模型, 在此基础上, 通过研究降雨与滑坡的关系, 计算不同预警等级的降雨量阈值, 构建出滑坡时空联合预警预报模型, 并对模型进行了实际案例验证[11]。董力豪选择随机森林算法、卷积神经网络算法、多层感知器神经网络算法三种人工智能算法构建区域滑坡预警模型, 实例校验结果显示, 卷积神经网络模型与深度神经网络模型适用于二级预警, 而随机森林模型更适合多级预警[12]。上述研究从影响因子选取、算法选择、模型优化等多角度开展了相应的研究, 为人工智能在地质灾害区域预警领域的应用奠定了坚实基础。

本文基于历史灾害数据、地质环境调查数据及诱发因素数据, 通过训练样本集构建、模型训练与优化评估等环节, 构建基于机器学习的滑坡灾害区域预警模型, 并进行实例校验。

2. 基于机器学习的滑坡灾害区域预警模型构建

2.1. 训练样本集构建

2.1.1. 影响因子选取

地质灾害的发生是多个内外因素综合作用的结果, 需要结合研究区的地质环境条件和地质灾害特征选取因子构建学习样本数据。通过对巴中市滑坡灾害形成条件的分析, 可以发现地形地貌、岩土体类型、

降雨模式以及人类工程活动是诱发滑坡的主要影响因素。具体而言，地形地貌的影响因子包括高程、坡度、坡向以及平面曲率；岩土体类型的影响因子则涉及地层岩性与岩土体结构特征；降雨模式主要包括长期和短期降雨；而人类工程活动主要指农业耕种和建设活动。

然而，由于岩土体结构特征、农业耕种和工程建设相关的大范围矢量数据难以获取，它们在本研究中并未被列为影响因子。因此，经过综合考量后，我们最终确定了以下影响因子：高程、坡度、坡向、平面曲率、地层岩性、长期降雨以及短期降雨。在这些因子中，高程、坡度、坡向、平面曲率以及地层岩性和长期降雨被视为静态因素，因为它们的值一旦确定，在研究期限内基本保持不变；而短期降雨则被视为动态因素，因为它的变化可能会在短时间内有较大变化，并对地质灾害的发生产生显著影响。

2.1.2. 数据来源

静态因子：高程、坡度、坡向、平面曲率均基于数字高程模型(DEM)提取，数据源为ALOS卫星采集的12.5 m分辨率DEM栅格数据。岩土体类别基于1:20万地质图提取，数据源为全国地质资料馆公开的中国地质调查局整理的全国1:200,000数字地质图(公开版)矢量数据。年均降雨量基于全国1 km分辨率年降雨数据提取，数据源为国家地球系统科学数据中心公开的西北农林科技大学团队整理的中国1 km分辨率年降水量栅格数据(1901~2023年)。

动态因子：短期降雨数据包括灾害发生当日降雨和前期有效降雨，其中前期有效降雨主要考虑灾害发生前3日的降雨量。两类数据类型没有区别，基于全国地面气象站基本观测数据提取，数据源为国家气象科学中心提供的全国5 km × 5 km逐日降水量栅格数据。

地质灾害点：收集了自然资源系统统计的1976年至2024年的1359处滑坡数据。

2.1.3. 数据处理

上述滑坡影响因子数据均为栅格类型，模型无法直接识别，需要先转换为点数据，采用相同的点文件将各因子数据提取至点文件中并生成表格，形成影响因子样本库。历史灾害点原始数据即为表数据，剔除无关信息后形成灾害点样本库。

数据分析：分别对灾害点和非灾害点的静态因素和动态因素进行对比分析，分析技术路线见图1。对于静态因素分析，通过绘制区域图，分析灾害点和非灾害点的静态因素的高频率取值；通过绘制散点图，分别找出灾害点和非灾害点的静态因素取值的区间段。对于动态因素分析，通过绘制当日降雨、前三日降雨、前三日降雨总和的热力分布图，分析不同降雨特征与灾害点的关联度。

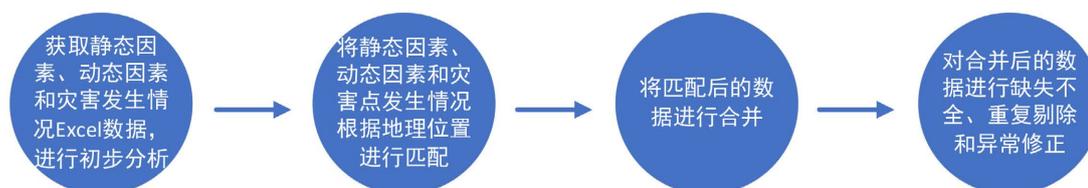


Figure 1. Data analysis technology process

图 1. 数据分析技术流程

数据对齐：将静态因子、降雨和灾害点数据按照经纬度进行对齐。其中灾害点和静态因子与降雨经纬度坐标并不一致，通过中心距离最近法进行匹配。

数据合并：将离散的静态因子、降雨和灾害点数据合并为一套统一的数据集合。

数据清洗：对数据集中的缺失数据、重复数据和异常数据进行处理。对于缺失数据，使用临近补全法处理；对于重复数据删除重复项；对于异常数据采用均值法处理。

2.2. 特征工程

根据区域预警 AI 模型的原始输入数据分析情况，对于高程、年均降雨量、平面曲率、坡度、坡向、岩土类别六项静态影响因子，对于预测区域内的每一个分割点来说，这些数据是稳定不变的。并且根据专家经验，这些因素和区域发生灾害的危险等级有直接的关联关系，因此这些因素可以直接作为模型所需的特征使用。

对于历史降雨和当日降雨等动态影响因子，将降雨对区域发生灾害的危险等级的影响分为即时性影响和滞后性影响。比如突降暴雨等当日降雨量大的情况，可能会在短时间内导致灾害的发生；而持续的降雨，可能会导致灾害的滞后发生。结合专家经验本文将当日降雨和前 3 日降雨之和作为特征使用。

2.3. 构造训练和测试数据

数据筛选：针对正负样本数据严重失衡的问题，对于负样本数据进行过滤和筛选，分别采用最近距离和最小因子法进行过滤，减少负样本数量，平衡正负样本比例[13]。

训练集和测试集生成：对于训练集，根据不同的筛选结果，选择发生日期、因素索引、灾害点索引、类别索引、经纬度、静态因子、当日降雨、前 3 日降雨之和、灾害点标签等数据项生成测试集。对于测试集，为每个灾害点计算当前及前 3 天的降雨数据，生成一组测试数据。

2.4. 模型与工具选择

人工神经网络包括：BP 神经网络、CNN 神经网络、RNN 神经网络等多种类别。CNN 神经网络模型比较适合特征值较多的输入，即图像识别等，因为特征值较多如果将神经元全连接会产生很大的计算负担所以通过卷积将特征值减少后再喂入神经网络，区域预警模型中的特征值数量在十个左右，其复杂度不需要采用卷积；RNN 模型适合处理时间序列的数据，区域预警模型中的大部分数据是静态因素，这些数据是根据地域分布的，不具有时间特性，不适合采用 RNN 模型。所以本文采用了具有全连接属性的 BP 神经网络。

本文模型训练过程是作者采用 python 软件实现，并调用 scikit-learn 机器学习库中的 MLPClassifier 分类算法完成模型本地训练。

2.5. 模型调参和训练

模型的训练和优化过程首先是确定输入、输出、学习率训练轮数等基本参数构建 BP 神经网络模型，然后再利用 hyperopt 方法实现神经网络中隐藏层层数、神经元个数、映射函数、优化器等超参数的优化选择[14]。

利用生成的多组测试集分别对训练完成的神经网络模型进行测试。通过绘制预测结果的热力分布图、预测灾害点的概率分布图，计算正负样本预测正确率、错误率等对多组模型进行分析，找到效果最佳的模型。

模型初始学习率设置为 0.001，训练到的轮数初始设置值为 1000，衰减的周期定义为 100，衰减率系数设置为 0.96，使用连续衰减[15]。经过不断尝试发现神经网络一般会在 3500 轮后下降较缓慢，于是将训练轮数设定在了 5000 轮。

优化后的模型具有一个输入层、两个隐藏层、一个输出层。每层节点数分别为 8、200、100、2。迭代次数 1877 次、最小损失值 0.16738，激活函数 relu，输出激活函数 logistic，寻优方法 sgd，学习率动态自适应。

2.6. 预警等级划分

区域预警模型实际应用中, 需要依据模型输出的灾害预测概率划分灾害预警等级, 参考地质灾害区域气象风险预警标准(试行)(T/CAGHP039-2018), 结合研究区具体情况, 将区域预警模型地质灾害预警等级划分标准确定为: 预测概率在区间[40%, 60%)内时为三级黄色预警; 预测概率在区间[60%, 90%)内时为二级橙色预警; 预测概率在区间[90%, 100%)内时为一级红色预警(表 1)。

Table 1. Early warning level division

表 1. 预警等级划分

预警等级	灾害发生可能性等级	概率 P
一级红色预警	可能性很高	[90%, 100%)
二级橙色预警	可能性高	[60%, 90%)
三级黄色预警	可能性较高	[40%, 60%)

3. 人工智能区域预警模型验证

3.1. 预警模型检验

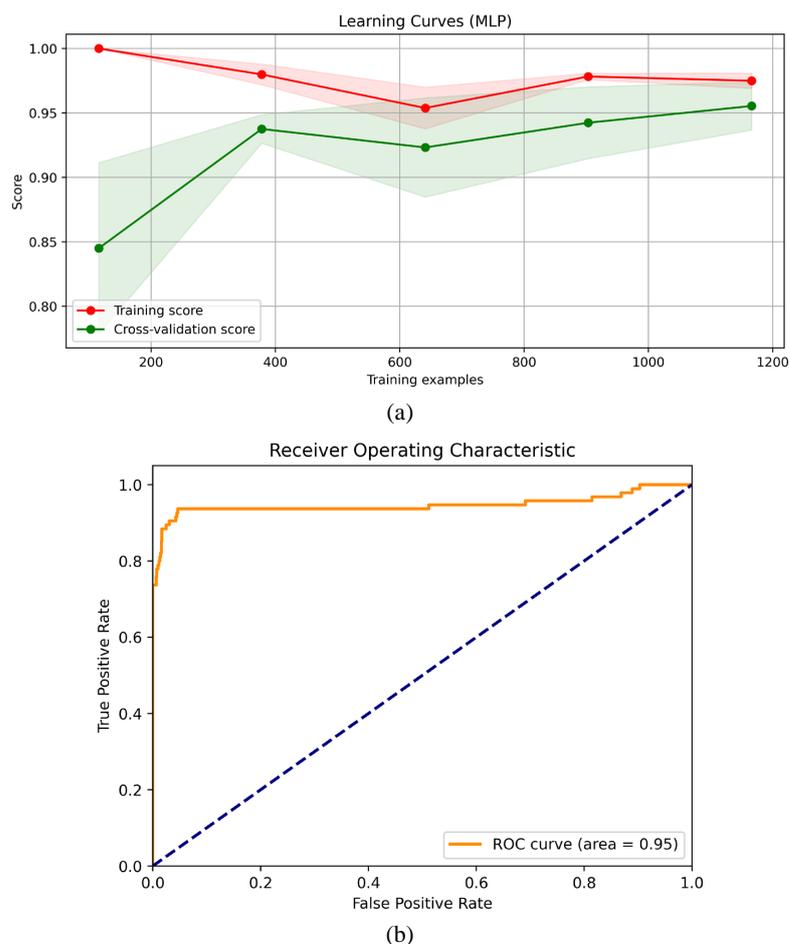


Figure 2. Learning curve and ROC curve of the model

图 2. 模型学习曲线和 ROC 曲线

采用准确率(Accuracy)、ROC 曲线和学习曲线(Learning Curve)来检验模型。使用测试集对模型的准确率和泛化能力指标进行评估,评估结果见图 2。模型准确率为 0.934, AUC 值为 0.95, 均较好, 模型不存在过拟合或欠拟合的情况(训练分数和交叉验证分数随着样本数量增多越来越接近, 且接近 1)。

3.2. 预警结果呈现效果对比

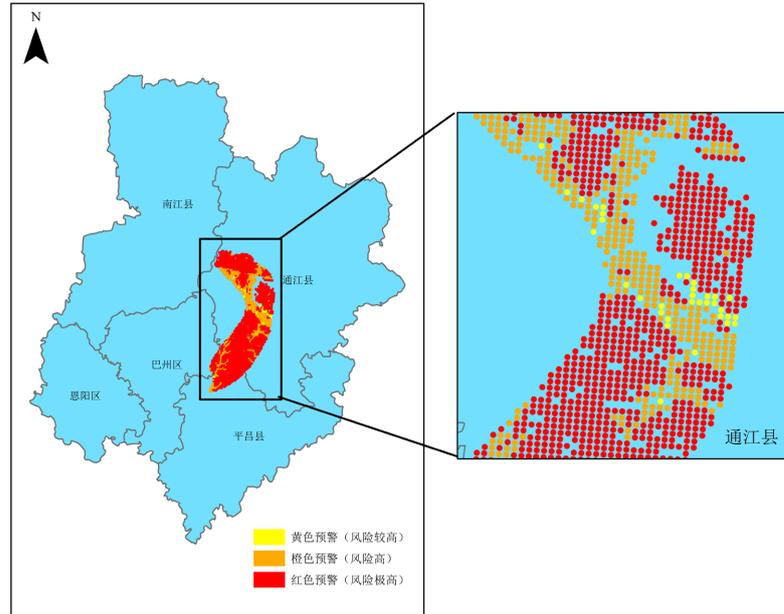


Figure 3. Warning results of AI regional warning model

图 3. AI 区域预警模型预警结果



Figure 4. Early warning results of statistical model

图 4. 统计模型预警结果

将 AI 区域预警模型预警结果与统计区域预警模型预警结果进行对比,其中统计区域预警结果采用的是省自然资源厅发布的预警结果。以 2023 年 5 月 30 日发布的预警结果为例(AI 区域预警模型预警结果见图 3,统计区域预警模型预警结果见图 4),可以看到 AI 模型结果的预警单元为 500 m×500 m 的栅格,涉及通江县西南部和平昌县北部部分区域,面积约 773 km²,而统计模型的预警单元为面状区域,涉及巴中市所有区县,面积约 15,705 km²。AI 模型的预警单元相对统计模型的预警单元更加精细,预警范围更小,对于灾害预警工作更具有指导意义。

3.3. 预警结果准确率对比

采用自然资源系统统计的 2023 年巴中市发生灾害进行模型准确率验证。2023 年巴中市共计发生 11 处小型滑坡,其中中型规模滑坡 2 处,其余为小型滑坡,中型险情等级滑坡 2 处,其余为小型险情,灾害点及验证信息见表 2。

成功预警情况也即预报准确率情况,如果在预警期内有地质灾害点落入预警区内,则表示此次预警准确。以 2023 年研究区内所有发生的灾害作为验证数据,AI 区域预警模型在 11 处灾害点中 7 处预报准确,准确率为 63%,其中两处中型险情灾害均成功预警;统计区域预警模型有 1 处预警准确,准确率仅为 9%。在 2023 年的预警应用中 AI 区域预警模型相较统计区域预警模型预警准确率有明显优势。

Table 2. Comparison of the accuracy of the AI regional early warning model and the statistical early warning model
表 2. 人工智能区域预警模型与统计预警模型准确率对比

灾害点地理位置	规模等级	险情等级	发生时间	经度	纬度	是否在预警区内	
						AI 模型	统计模型
奇章街道石燕社区 2 组	小型	小型	2023/3/27	106.926	31.860	是	否
奇章街道郑家山村 2 组	小型	中型	2023/3/27	106.908	31.865	是	否
元山镇插旗山村 11 组	中型	中型	2023/5/15	107.173	31.710	是	否
壁州街道西郊村 1 组	小型	小型	2023/5/21	107.227	31.919	是	否
化成镇长滩河村 1 组	小型	小型	2023/5/31	106.921	31.938	是	是
得胜镇红星村 1 组	小型	小型	2023/6/10	107.161	31.824	否	否
玉堂街道侯家沟村 3 组	小型	小型	2023/7/26	106.853	31.898	否	否
涪阳镇红江村 5 组	小型	小型	2023/9/25	107.170	32.088	否	否
新场镇毛村 5 组	中型	小型	2023/10/5	107.137	32.158	否	否
诺江镇千佛村 4 组	小型	小型	2023/10/26	107.247	31.941	是	否
枣林镇八字村 1 组	小型	小型	2023/12/25	106.730	31.920	是	否
预警准确率	(预警区内灾害点数量/灾害点总数量)					63%	9%

4. 结论

本文基于人工智能技术构建了滑坡灾害区域预警模型,采用准确率、ROC 曲线和学习曲线来检验模型,结果显示,模型准确率为 0.934, AUC 值为 0.95,模型准确率及泛化能力均较好。同时通过实例进行了预警检验,根据结果对比,基于人工智能的地质灾害区域预警模型相较统计预警模型的预警结果准确率有明显改善,预警结果的刻画也要更加精细,对滑坡灾害区域风险精细化防控有较好的指导作用。

参考文献

- [1] 刘传正, 温铭生, 唐灿. 中国地质灾害气象预警初步研究[J]. 地质通报, 2004, 23(4): 303-309.
- [2] 刘传正, 刘艳辉. 地质灾害区域预警原理与显式预警系统设计研究[J]. 水文地质工程地质, 2007, 34(6): 109-115, 125.
- [3] 刘传正, 刘艳辉, 温铭生, 等. 中国地质灾害区域预警方法与应用[M]. 北京: 地质出版社, 2009.
- [4] 杨寅, 林建, 包红军. 2018 年地质灾害重点区域气象预警效果检验[J]. 中国地质灾害与防治学报, 2019, 30(6): 135-140.
- [5] 刘艳辉, 苏永超. 四川青川县区域地质灾害气象风险预警模型研究[J]. 工程地质学报, 2019, 27(1): 134-143.
- [6] 刘艳辉, 方然可, 苏永超, 等. 基于机器学习的区域滑坡灾害预警模型研究[J]. 工程地质学报, 2021, 29(1): 116-124.
- [7] 李阳春, 刘黔云, 李潇. 基于机器学习的滑坡崩塌地质灾害气象风险预警研究[J]. 中国地质灾害与防治学报, 2021, 32(3): 118-123.
- [8] 刘艳辉, 黄俊宝, 肖锐铧, 等. 基于随机森林的福建省区域滑坡灾害预警模型研究[J]. 工程地质学报, 2022, 30(3): 944-955.
- [9] 谢旭阳, 江田汉, 王云海, 等. 基于支持向量机的尾矿库灾害区域预警[J]. 中国安全生产科学技术, 2008, 4(4): 17-21.
- [10] 黄桥, 杨洋, 吴思. 基于机器学习方法的长输管线地质灾害预测建模[J]. 测绘与空间地理信息, 2019, 42(11): 97-100.
- [11] 孙德亮. 基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D]: [博士学位论文]. 上海: 华东师范大学, 2019.
- [12] 董力豪. 基于人工智能算法的福建省区域滑坡预警模型的研究[D]: [硕士学位论文]. 郑州: 华北水利水电大学, 2023.
- [13] 郭衍昊, 窦杰, 向子林, 等. 基于优化负样本采样策略的梯度提升决策树与随机森林的汶川同震滑坡易发性评价[J]. 地质科技通报, 2024, 43(3): 251-265.
- [14] 全卫国, 李敏霞, 张一可. 深度学习优化算法研究[J]. 计算机科学, 2018, 45(S2): 155-159.
- [15] 沈花玉, 王兆霞, 高成耀, 等. BP 神经网络隐含层单元数的确定[J]. 天津理工大学学报, 2008, 24(5): 13-15.