



基于机器学习的地质灾害易发性研究

——以山东省平邑县为例

高洪军¹, 卞宝文^{2*}, 王欣瑶²

(1.日照市自然资源和规划局, 山东 日照 276800; 2.山东省煤田地质局第一勘探队, 山东 青岛 266000)

摘要:地质灾害的早期识别和易发生区域的监测是防灾减灾的重要工作。本文以山东省平邑县为研究区域, 将 GF-1 WFV 光学影像、ASTER GDEM 地形数据和降水数据融合为多源异构数据, 对比了 TensorFlow 算法、支持向量机和随机森林 3 种机器学习算法对地质灾害易发性区域的提取效果, 提取了研究区 2021—2024 年同时期地质灾害易发性区域, 认为 TensorFlow 算法、支持向量机和随机森林均能够较好的识别是滑坡易发生区域, 其中 TensorFlow 算法相较于其他方法的分类精度较高, 总体精度为 82.33%, Kappa 系数为 0.82。2021—2024 年, 平邑县易发生地质灾害的区域面积占比为 11.5%~12.5%, 主要集中在研究区西北的蒙山大洼区、唐村水库南部和九间棚区域。研究成果可为地质灾害易发性区域提取算法的选择和地质灾害预防提供参考。

关键词:地质灾害易发性提取; 机器学习; 多源数据; 山东平邑

中图分类号: P237

文献标识码: A

doi: 10.12128/j.issn.1672-6979.2025.07.005

0 引言

地质灾害的早期识别和易发生区域的监测是防灾减灾的重要工作。山东省的地貌特征呈现为山高谷深、崖陡壁峭, 且不同岩性组合和强烈的地质构造运动加剧了该区域地质结构的不稳定性。而季风区域大陆性气候, 雨量集中, 灾害性天气频发, 使鲁中南地区成为地质灾害易发生区域^[1-4]。本文选择山东省临沂市平邑县作为研究区域, 进行地质灾害易发性区域的识别和提取。

由于遥感技术的快速发展, 利用遥感影像对地质灾害早期识别和易发生区域的监测已成为防灾减灾的重要手段。多光谱影像结合了地形数据和气候数据的多源异构数据, 包含了更加丰富的特征信息, 有助于地质灾害易发性区域的提取。如裘湧泉^[5]基于高分二号和哨兵 2A 影像, 根据构建的识别规则, 利用决策树分类的面向对象方法提取滑坡信息, 正确提取率为 82.40%; 白石^[6]等融合高分辨率遥感影像和数字高程模型数据, 利用面向对象多特征变化向量分析法, 对研究区内的滑坡区域进行识别。多

源异构数据丰富的特征信息, 也为利用优化的机器学习算法进行地质灾害易发性区域的提取提供了支撑。近年来, 通过机器学习算法对遥感影像进行解译, 从而进行地质灾害的早期识别和易发性研究取得了众多的研究成果^[7]。如 Stumpf^[8]等人提出了一种结合面向对象图像分析技术和随机森林算法的监督分类方法, 该分类方法能够减少人工标注工作, 并以客观的方式选取重要的对象特征和分类阈值, 提高识别过程的客观性和效率; 胡文杰^[9]以 Swin Transformer Block 为基础构建设计集成了一种交叉注意力机制模块, 提升了滑坡的识别率; 岳帅帅^[10]通过基于随机森林的面向对象分类提取到略阳县内 505 处滑坡体, 总体精度为 86.3%, 展现出基于随机森林的面向对象分类法在识别滑坡中较高的精度和适应性。

在地质灾害易发性区域提取领域, 机器学习算法已展现出显著的成效。这些算法具备自动化识别地质灾害易发性区域的能力, 能够高效地从多源异构数据中提取关键特征信息, 进而实现对地质灾害易发性区域的精确界定。此外, 机器学习算法在计

收稿日期: 2025-01-09; **修订日期:** 2025-03-27; **编辑:** 武昕普

作者简介: 高洪军(1980—), 男, 山东日照人, 高级工程师, 主要从事自然资源管理和应用工作; E-mail: ghj-0210@163.com

* **通信作者:** 卞宝文(1990—), 男, 山东日照人, 工程师, 主要从事测绘工程相关工作; E-mail: 173308946@qq.com

算成本方面相对较低,且通常能够达到较高的识别精度,足以满足多样化的研究和实际应用需求。基于此,本文将 GF-1 WFV 光学影像、ASTER GDEM 地形数据和降水数据进行数据处理,形成多源异构数据,通过对比 TensorFlow 算法、支持向量机和随机森林三种机器学习算法对地质灾害易发性区域的提取效果,确定适用于地质灾害易发性区域提取的优秀算法,为平邑县地质灾害预防提供参考。

1 研究区与数据

1.1 研究区概况

平邑县地处山东省临沂市西部、山东省东南部,县域总面积约为 1 825 km²。地质构造复杂,具有明显的山区特征,浚河、温凉河穿县而过,县内存在大量岩溶塌陷区域^[11]。最高点位于蒙山珠峰龟蒙顶,海拔约为 1 155 m,最低点在县北部孙家庄区域,海拔约为 97 m。全境地势南北高,中间低,略向东南倾斜。平邑县属暖温带东南亚季风区域大陆性气候,寒暑交替、四季明显,雨量集中,旱涝不均^[12]。夏季雨量集中,灾害性天气频繁,易引发山体滑坡、泥石流等地质灾害。

1.2 数据及预处理

(1)光学遥感数据。云量筛选了高分 1 号 2021—2024 年 4 景遥感数据。高分一号卫星(GF-1)搭载了 2 台 2 m 分辨率全色、8 m 分辨率多光谱相机和 4 台 16 m 分辨率多光谱相机,本次使用 16 m 分辨率多光谱宽幅影像。GF-1 WFV 获取的 16 m 分辨率能够为自然资源部门、地质部门提供高精度、宽范围的空间观测服务^[13]。所有光学遥感数据均经过辐射定标、大气校正和正射校正,为确保多源异构数据在实验中保持空间分辨率一致,对所有数据均重采样至 30 m 分辨率^[14]。

(2)地形数据。本次使用的地形数据是先进星载热发射和反射辐射仪全球数字高程模型(ASTER GDEM),该数据是由日本 METI 和美国 NASA 联合研制并免费面向公众分发^[15]。该数据是根据 NASA 的新一代对地观测卫星 Terra 的详尽观测结果制作完成,其数据覆盖范围为 83°N 到 83°S 之间的所有陆地区域,达到了地球陆地表面的 99 %^[16]。

(3)降水数据。本文根据平邑县发布的各镇街年度累计降雨量,以各镇街的地理中心作为降雨数据点,通过克里金插值方法,得到了平邑县年度累计降雨量图。数据的具体情况见表 1。

表 1 数据使用具体情况

| 多元异构数据 | 数据来源 | 空间分辨率 /m | 数据详情 |
|--------|---------------------------------------|----------|--|
| 光学遥感数据 | GF-1 WFV 卫星影像 | 16 | GF1_WFV2_E117.5_N36.0_20211127_L1A0006086676 GF1_WFV2_E118.2_N36.0_20221205_L1A0006968055 GF1_WFV4_E118.2_N36.0_20231222_L1A13214287001 GF1_WFV4_E117.5_N36.0_20241221_L1A13689573001 |
| 地形数据 | 先进星载热发射和反射辐射仪 全球数字高程模型(ASTER GDEM) | 30 | ASTGTMV003_N35E117_dem |
| 降水数据 | 平邑县各镇街年度累计降雨量数据 | 30 | 克里金插值 |

2 研究方法

2.1 地质灾害易发性识别方法

目前,在基于多源遥感数据进行地质灾害易发性识别的研究中,已经有包括支持向量机、随机森林和 TensorFlow 算法等机器学习算法。为寻找适用于研究区地质灾害易发性识别的算法,本文对比了支持向量机、随机森林、TensorFlow 算法^[17]。

(1)TensorFlow 算法。TensorFlow 是由 Google 团队开发的一款用于机器学习研究和产品开发的深度学习框架,拥有高度灵活性和可扩展性,

常用于开发自然语言处理、图像识别、手写识别等各种任务模型^[18]。其中,Tensor 表示张量,即 N 维数组,Flow 表示数据流图,直观表达了张量之间通过计算相互转化的过程,高度的灵活性使得 TensorFlow 成为主流的遥感影像识别和分析工具^[19]。

(2)支持向量机是一种源自统计学习理论的监督分类方法,使用被称为最优决策超平面的面将类分开,使类之间的边界最大化,最接近最优决策超平面的数据点称为支持向量。同时,在使用支持向量机方法进行分类时,需要选择一个核函数,用于在估计目标类时提供附近数据点的权重^[20]。

(3)随机森林算法由 L.Breiman 提出,该算法通

过构建大量不相关的随机决策树,并在使用多数投票决策的方式对所有生成的决策树进行引导和聚合,从而实现对图像的分类^[21]。

2.2 精度验证

为确定滑坡易发区域的提取精度,在研究区内随机采集 300 个样本点,根据天地图影像进行目视判读,确定检验样本点所在区域是否为易发区域,以此判断提取结果的精度。精度验证采用混淆矩阵和 Kappa 系数作为指标。通过混淆矩阵可以得到总体精度(Overall Accuracy, OA)、用户精度(User Ac-

curacy, UA)和生产者精度(Producer Accuracy, PA)等一系列的精度评价指标。而 Kappa 系数被广泛应用于遥感分类的精度评价^[22]。

3 结果与讨论

3.1 识别精度比较

通过使用 3 种机器学习算法,对研究区进行分类,共分为水体、林地、农业用地、建筑用地、草地和地质灾害易发性区域 6 类地类,具体分类结果对比如图 1 所示。

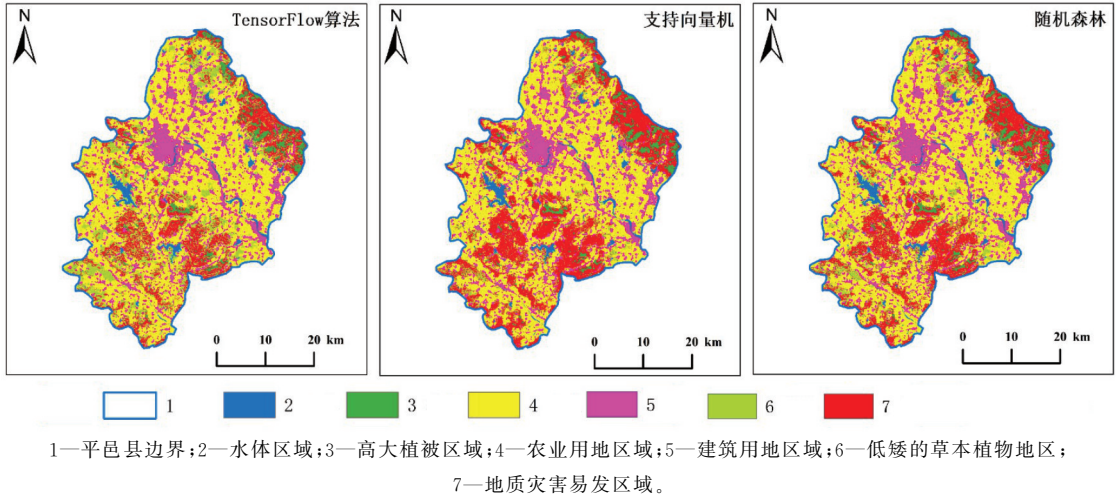


图 1 各算法分类结果

由各算法分类结果图对比可以发现,各算法对于水体、建筑用地和林地的分类结果较为一致,地质灾害易发性区域主要集中在研究区西北的蒙山大洼区、唐村水库南部和九间棚区域。相较于支持向量机和随机森林的分类结果,TensorFlow 算法的分类结果中易发性区域面积明显更小,且草地面积更大,更加符合研究区地质环境特征^[23]。

为进一步确认各算法的分类精度,本文在研究区内随机采集 300 个检验样本点,根据天地图高分辨率的影像对样本点所属区域的正确类别进行目视解译。通过精度验证,得到 3 种机器学习算法对地质灾害易发性区域和各类地物的识别精度和 Kappa 系数,具体结果如表 2 所示。

不同分类方法的总体精度均不小于 75%,Kappa 系数不小于 0.75。其中 TensorFlow 算法的分类结果总体精度最高,为 82.33%,Kappa 系数为 0.82;随机森林的分类精度次之,支持向量机的分类精度

表 2 提取精度评价指标

| 地类 | TensorFlow 算法 | | 支持向量机 | | 随机森林 | |
|----------|---------------|-------|-------|-------|-------|-------|
| | UA/% | PA/% | UA/% | PA/% | UA/% | PA/% |
| 水体 | 84.00 | 84.00 | 78.26 | 75.00 | 82.98 | 81.25 |
| 林地 | 81.63 | 88.89 | 80.85 | 86.36 | 80.85 | 86.36 |
| 农业用地 | 78.00 | 73.58 | 64.91 | 67.27 | 69.09 | 67.86 |
| 建筑用地 | 94.00 | 82.46 | 93.48 | 79.63 | 93.75 | 78.95 |
| 草地 | 77.08 | 80.43 | 66.67 | 61.54 | 70.21 | 68.75 |
| 地质灾害易发区 | 79.25 | 85.71 | 69.64 | 82.98 | 71.43 | 85.11 |
| OA | 82.33 | | 75.00 | | 77.67 | |
| Kappa 系数 | 82 | | 75 | | 78 | |

最低。在对各类方法的分类结果进行比较后发现,支持向量机方法在小样本分类中更具有优势,但由于少量支持向量影响了精度,因此造成农业用地与草地这种易混淆的地物分类精度较差。随机森林方法的分类精度相对较高,但存在参数较难调优的缺点,同时其运行速度比另外两种分类算法更慢。总体而言,TensorFlow 算法相较于其他方法的分类精度较高,灵活性高,支持多种优化算法,可根据输入多源数据的特点进行参数调优,可以通过数据并行

和模型并行实现高效计算,非常适合多源遥感数据分析,因此本文使用 TensorFlow 算法进行平邑县地质灾害易发性区域的提取。

TensorFlow 算法提取结果中建筑用地的提取精度最高,用户精度为 94.00%,生产者精度为 82.46%。这是由于建筑用地在遥感影像上的特征较为明显,易于识别和提取。水体和林地的提取精度也相对较高,用户精度和生产者精度均在 80%以上。相比之下,农业用地和草地的提取精度稍低,这可能是由于农业用地和草地在遥感影像上的特征相似,且易受季节、气候等因素的影响,导致分类过程中存在一定的困难。地灾害易发区的用户精度为 79.25%,生产者精度为 85.71%,相对其他地类而言略低,但考虑到地灾害易发区在总体面积中的占比可能较小,且特征复杂,这样的精度表现仍可接受。此外,其用户精度和生产者精度的差异也表明,在分

类过程中可能存在一定程度的误判和漏判。综上所述,TensorFlow 算法在不同地类上的提取精度表现出一定的差异性,但整体而言,提取结果的精度较高,其分类结果有较高的参考价值。

3.2 时空变化分析

通过 TensorFlow 算法对 2021—2024 年平邑县同时期影像进行地质灾害易发区域提取,地质灾害易发性区域主要集中在研究区西北的蒙山大洼区、唐村水库南部和九间棚区域。2023 年提取结果中九间棚区域和唐村水库南部的地质灾害易发性区域面积有所增加。就其他地类而言,浚河、温凉河两条河流贯穿研究区,为主要水系,建筑面积在现有的位置逐步外扩,草地区域主要集中在研究区的南部,与农业用地交错,林地主要集中要蒙山的大洼区和龟蒙区(图 2)。

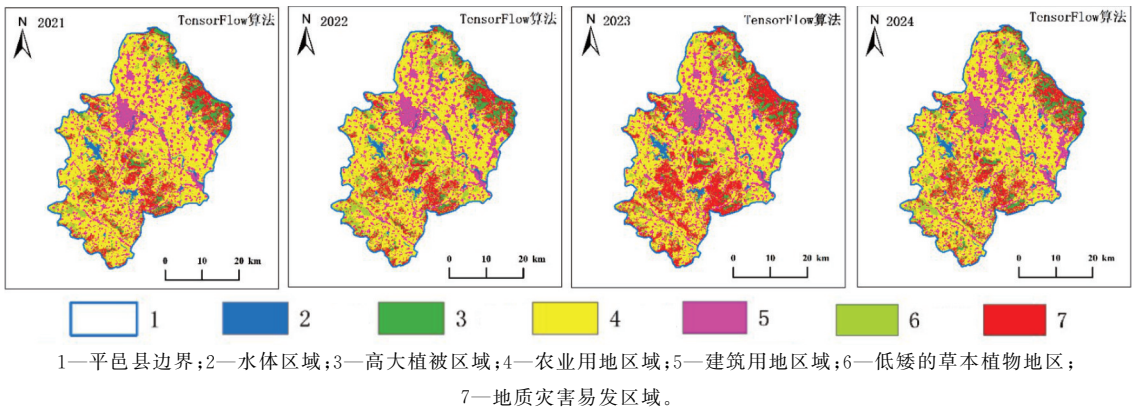


图 2 地质灾害易发性区域识别

为验证数据的准确性,将 TensorFlow 算法提取的各类型土地利用类型的面积与公开发布的全球土地利用数据进行对比,该分类结果与 Sentinel-2 世界 10 m 土地利用时间序列数据具有高度的一致性^[24]。表 3 详细体现了 2021—2024 年期间平邑县地质灾害易发性区域面积的变化,总体上呈现出先稳定后增长再减少的趋势。从 2021 年的 227.211 km²到 2022 年的 226.384 km²,面积略有减少,但占比保持稳定,分别为 12.46%和 12.41%。然而,到了 2023 年,地灾易发区面积显著增加至 254.460 km²,占比也相应提升至 13.95%。这种增长可能与气候变化、降水增加、人类活动对地质环境的干扰等因素有关。值得注意的是,尽管 2024 年地灾易发区面积有所减少至 212.868 km²,但其在总

面积中的占比仍然较高,达到 11.67%,地质灾害的防范和治理工作依然面临严峻挑战。

| 表 3 地质灾害易发性区域提取结果 | | |
|-------------------|--------------------|-------|
| 年份 | 面积/km ² | 占比/% |
| 2021 年 | 227.211 | 12.46 |
| 2022 年 | 226.384 | 12.41 |
| 2023 年 | 254.46 | 13.95 |
| 2024 年 | 212.868 | 11.67 |

4 结论

本文以山东省平邑县为研究区域,将 GF-1 WFV 光学影像、ASTER GDEM 地形数据和降水数据进行数据处理,形成多源异构数据,通过对比 TensorFlow 算法、支持向量机和随机森林 3 种机器学习算法对地质灾害易发性区域的提取效果,在确

定算法后提取 2021 年和 2024 年同时期地质灾害易发性区域,得到以下结论:

(1)TensorFlow 算法、支持向量机和随机森林均能够较好地识别滑坡易发生区域,总体精度均不小于 75%,Kappa 系数均不小于 0.75。

(2)TensorFlow 算法相较于其他方法的分类精度较高,总体精度为 82.33%,Kappa 系数为 0.82,该算法可根据输入多源数据的特点进行参数调优,适用于进行地质灾害易发性区域的提取。

(3)2021—2024 年,平邑县易发生地质灾害的区域面积占比为 11.5%~12.5%,相较于 2021 年,2024 年滑坡易发生区域面积减少了 14.343 km²,地质灾害易发性区域主要集中在研究区西北的蒙山大洼区、唐村水库南部和九间棚区域。

本文对比了 TensorFlow 算法、支持向量机和随机森林 3 种机器学习算法对地质灾害易发性区域的提取效果,未通过标准样本对其他算法进行提取效果的对比,如何利用多源异构数据的优势,优化各类机器学习算法的提取精度和效率有待深入研究。

参考文献:

[1] 辛鲁斌,韩玲,李良志.基于多源数据融合的滑坡智能识别[J].地球科学与环境学报,2023,45(4):920-928.

[2] 孙萍萍,张茂省,贾俊,等.中国西部黄土区地质灾害调查研究进展[J].西北地质,2022(3):55.

[3] 高峰,孟凡奇,张丽,等.山东省地质灾害调查工作回顾与展望[J].山东国土资源,2022,38(10):35-41.

[4] 常允新,宋长斌.山东省崩塌、滑坡、泥石流灾害发育现状与趋势分析[J].山东国土资源,2000,16(4):39-44.

[5] 裘湧泉.基于多源遥感数据的黄土滑坡提取[D].江西:东华理工大学,2023.

[6] 白石,唐攀攀,苗朝,等.基于高分辨率遥感影像和改进 U-Net 模型的滑坡提取:以汶川地区为例[J].自然资源遥感,2024,36(3):96-107.

[7] 韩学林,王秀芬,郭宝奎,等.山东省砂石资源遥感解译成果及应用[J].山东国土资源,2023,39(11):59-67.

[8] Stumpf A, Kerle N. Object-oriented mapping of landslides using Random Forests[J]. Remote sensing of environment,

2011, 115(10): 2564-2577.

[9] 胡文杰.基于深度学习的滑坡识别方法:以四川省泸定县为例[D].廊坊:防灾科技学院,2024.

[10] 岳帅帅.基于多源遥感数据的滑坡识别技术研究[D].西安:长安大学,2023.

[11] 孙硕.岩溶塌陷治理设计与实现:以山东省平邑县泰和村岩溶塌陷设计治理点为例[J].山东国土资源,2023,39(10):26-30.

[12] 冯广,杜现福,孙超,等.山东省平邑县九间棚村土壤地球化学调查评价[J].山东国土资源,2024,40(10):35-42.

[13] 马艳敏,袁福香,李建平,等.基于 GF-1 卫星数据的敦化市园参分布提取与变化分析[J].气象灾害防御,2024,31(3):45-48.

[14] 袁金国,牛铮,王锡平.基于 FLAASH 的 Hyperion 高光谱影像大气校正[J].光谱学与光谱分析,2009,29(5):1181-1185.

[15] 张朝忙,刘庆生,刘高焕,等.SRTM3 与 ASTERGDEM 数据处理及应用进展[J].地理与地理信息科学,2012,28(5):29-34.

[16] 卫石印,李忠涵,张启慧,等.基于多源遥感数据的滑坡易发区域识别:以山东省平邑县为例[J].山东国土资源,2024,40(10):60-65.

[17] 王家柱,铁永波,白永健,等.机器学习在斜坡地质灾害领域的应用现状与展望[J/OL].水文地质工程地质,1-13.[2025-03-27].DOI:10.16030/j.cnki.issn.1000-3665.202402011.

[18] 康恺,何伟,董天祯,等.一种基于 TensorFlow 的蔬菜识别技术[J].中国科技信息,2023(23):103-106.

[19] 荣光旭,李宗洋,田凯.TensorFlow 多元线性回归在边坡稳定性分析中的应用[J].齐鲁工业大学学报,2020,34(6):6.

[20] HSU C W, CHAN C C, LIN C J. A Practical Guide to Support Vector Classification[J].BJU International, 2008, 101(1):1396-1400.

[21] 黄衍,查伟雄.随机森林与支持向量机分类性能比较[J].软件,2012,33(6):4.

[22] 朱江涛.基于 GEE 的鄱阳湖湿地植被长期变化特征及影响因素研究[D].南昌:东华理工大学,2022.

[23] 尹常铭.基于机器学习的滑坡易发性区划和滑坡快速识别研究[D].大连:大连理工大学,2022.

[24] Karra K, Kontgis C, Statman-Weil Z, et al. Global land use / land cover with Sentinel 2 and deep learning[C]// 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021:4704-4707.

Study on the Susceptibility of Geological Hazards
Based on Machine Learning

——Taking Pingyi County in Shandong Province as an Example

GAO Hongjun¹, BIAN Baowen², WANG Xinyao²

(1. Rizhao Bureau of Natural Resources and Planning, Shandong Rizhao 276800, China; 2.No.1 Exploration Brigade of Shandong Coalfield Geologic Bureau, Shandong Qingdao 266000, China)

Abstract: Early identification of geological disasters and monitoring of easy—happening areas are important work in disaster prevention and reduction. In this paper, taking Pingyi county in Shandong province as the study area, the GF - 1 WFV optical image, ASTER GDEM terrain data and precipitation data are fused into multi—source heterogeneous data. The extraction effects of three machine learning algorithms, such as TensorFlow algorithm, support vector machine, and random forest in geological hazard easy—happening areas have been compared. Geological hazard easy—happening areas in the study area from 2021 to 2024 have been extracted. By using TensorFlow algorithm, support vector machine and random forest methods, landslide easy—happening areas can all identified well. compared to other methods, TensorFlow algorithm has a higher classification accuracy with an overall accuracy of 82.33% and a Kappa coefficient of 0.82. From 2021 to 2024, the proportion of geological hazard easy—happening areas in Pingyi county ranged from 11.5% to 12.5%. The fluctuations are mainly concentrated in Mengshan Dawa area in the northwest of the study area, the southern part of Tangcun reservoir, and Jiu Jianpeng area. The research results can provide some references for the selection of extraction algorithms for geological hazard easy—happening areas and the prevention of geological hazards in Pingyi county in Shandong province.

Key words: Geological hazard susceptibility extraction; machine learning; multi—source data; Pingyi county in Shandong province