农业环境科学学报
JOURNAL OF AGRO-ENVIRONMENT SCIENCE
中文核心期刊/CSCD
请通过网上投稿系统投稿
网址：http://www.aes.org.cn

机器学习在生态环境损害鉴定评估领域的应用前景

武子豪, 吴礼滨, 洪伟, 丁泽聪, 易皓, 张晓园, 曾子龙, 崔恺

引用本文:
武子豪, 吴礼滨, 洪伟, 丁泽聪, 易皓, 张晓园, 曾子龙, 崔恺. 机器学习在生态环境损害鉴定评估领域的应用前景[J]. 农业环境科学学报, 2023, 42(12): 2860–2868.

在线阅读 View online: https://doi.org/10.11654/jaes.2023–0898

您可能感兴趣的其他文章

Articles you may be interested in

环境DNA技术在水生态领域应用研究进展
赵彦伟, 陈家琪, 董丽, 麻晓梅, 白洁, 田凯
农业环境科学学报. 2021, 40(10): 2057–2065 https://doi.org/10.11654/jaes.2021–0666

微塑料的环境行为及其生态毒性研究进展
刘沙沙, 付建平, 郭楚玲, 党志
农业环境科学学报. 2019, 38(5): 957–969 https://doi.org/10.11654/jaes.2018–1016

粤北大宝山矿区污染成因与源头控制技术应用进展
党志, 姚谦, 陈锴, 郑雄开, 付浩健, 廖正家, 卢桂宁
农业环境科学学报. 2021, 40(7): 1377–1386 https://doi.org/10.11654/jaes.2021–0610

退化喀斯特生态恢复过程对土壤微节肢动物群落多样性的影响
李宜蓉, 龙健, 李娟, 刘灵飞, 廖洪凯, 王显, 杨睿
农业环境科学学报. 2021, 40(2): 310–318 https://doi.org/10.11654/jaes.2020–0728

中国畜牧业环境约束的量化分析——基于土地环境承载力和生态足迹
熊学振, 杨春, 于琳
农业环境科学学报. 2021, 40(8): 1799–1807 https://doi.org/10.11654/jaes.2021–0136

关注微信公众号，获得更多资讯信息

武子豪，吴礼滨，洪伟，等. 机器学习在生态环境损害鉴定评估领域的应用前景[J]. 农业环境科学学报, 2023, 42(12): 2860-2868.
WU Z H, WU L B, HONG W, et al. Prospects of machine learning in the field of ecological environmental damage identification and assessment[J]. *Journal of Agro-Environment Science*, 2023, 42(12): 2860-2868.

开放科学OSID

机器学习在生态环境损害鉴定评估领域的应用前景

武子豪，吴礼滨，洪伟，丁泽聪，易皓，张晓园，曾子龙，崔恺*

（生态环境部华南环境科学研究所 华南环境损害司法鉴定中心，广州 510655）

摘　要：生态环境损害鉴定评估工作是环境行政处罚的重要依据，随着近年来生态环境损害案件的不断发生，损害鉴定工作流程复杂、案件资料分析工作量大、数据缺失严重等问题不断显现，如现场勘察耗时耗力、污染物溯源困难、基线不清、损害赔偿金额难以确定等。为解决这些问题，本文探讨了机器学习在生态环境损害鉴定评估中的应用前景。近几年，机器学习凭借其强大的计算能力，已在数据挖掘、图像识别和自然语言处理等领域发挥了重要作用，通过综述机器学习在上述领域的已有进展，结合生态环境损害鉴定评估总体工作流程，本文深入探究了机器学习在损害鉴定评估中的应用前景，分析了机器学习在鉴定评估工作中的挑战和局限性，指明机器学习的应用可以提高损害鉴定评估的工作效率，促进损害鉴定评估有序化、系统化发展。

关键词：生态环境损害鉴定评估；机器学习；图像识别；自然语言处理

中图分类号：X321　　**文献标志码：**A　　**文章编号：**1672-2043(2023)12-2860-09　　doi:10.11654/jaes.2023-0898

Prospects of machine learning in the field of ecological environmental damage identification and assessment

WU Zihao, WU Libin, HONG Wei, DING Zecong, YI Hao, ZHANG Xiaoyuan, ZENG Zilong, CUI Kai*

(South China Environmental Forensic Center, South China Institute of Environmental Science, Ministry of Ecology and Environment, Guangzhou 510655, China)

Abstract: Ecological environment damage appraisal and assessment is essential for environmental administrative punishment. With the continuous occurrence of ecological environment damage cases in recent years, the complexity of damage appraisal workflow, the workload of analyzing case information, and the seriousness of the data missing problems, such as time-consuming and exhausting *on-site* investigation, difficulty in the traceability of pollutants, unclear baseline, and difficulty in determining the damage compensation, continue to emerge. This study explored the prospects for applying machine learning in the appraisal and evaluation of ecological environmental damage to address these problems. Machine learning has been crucial in data mining, image recognition, and natural language processing in recent years through its powerful computational ability. By reviewing the existing progress of machine learning in the above fields, combined with the overall workflow of ecological environment damage appraisal with an in-depth exploration of the prospects for applying machine learning in damage appraisal, this study analyzed the challenges and limitations of applying machine learning in appraisals and indicates that it is challenging to solve these problems by its application. It also indicates that machine learning can improve the efficiency of damage appraisal and promote its orderly and systematicly development.

Keywords: ecological damage assessment; machine learning; image recognition; natural language processing

收稿日期：2023-10-30　　**录用日期：**2023-11-16

作者简介：武子豪(1997—)，男，广东珠海人，硕士，从事生态环境损害鉴定评估研究。E-mail：2535326734@qq.com

***通信作者：**崔恺　E-mail：cuikai@scies.org

基金项目：中央级公益性科研院所基本科研业务费专项资金项目(PM-zx703-202204-070，PM-zx703-202305-270，PM-zx703-202305-189)

Project supported: Fundamental Research Funds for the Central Public Welfare Research Institutes (PM-zx703-202204-070, PM-zx703-202305-270, PM-zx703-202305-189)

随着生态环境损害赔偿制度与环境公益诉讼制度的相继建立，我国进入生态环境损害鉴定评估工作高速发展时代，各地区发生的生态环境损害类案件不断增多，其鉴定评估工作成为环境司法审判的热点及难点问题。近几年，随着ChatGPT的横空出世，越来越多研究学者开始关注机器学习在各领域中的应用前景。机器学习在数据挖掘、图像处理、文本处理领域的应用可以在生态环境损害鉴定评估工作中的各个环节中发挥关键辅助作用，有助于鉴定评估工作的有序化、系统化和高质量发展。

1 生态环境损害鉴定评估介绍

目前国家高度重视生态环境保护工作，生态环境部门与司法部门联手强力打击环境污染和生态破坏犯罪，环境公益诉讼与环境损害赔偿制度构建愈发完善，生态环境损害鉴定评估作为技术支撑工作[1]，评估结果的质量高低对环境污染和生态破坏类案件的走向起到至关重要的作用。生态环境损害鉴定评估工作是指鉴定评估机构按照规定的程序和方法，运用科学技术和专业知识，鉴别污染物性质、评估环境污染或生态破坏行为所导致的环境损害范围及程度、判定环境污染或生态破坏行为与环境损害间的因果关系、确定生态环境恢复的基线水平、量化损害赔偿金额等的活动[2]。鉴定人在实际工作中往往会面临许多问题，如前期案件资料收集整理困难、难以提取关键有效信息、现场勘察工作耗时耗力、污染物溯源难、环境损害量化方式选取难、损害赔偿金额易受到外部质疑与挑战[3]等。因此，探索环境损害鉴定评估的高效、高质量工作方式成了目前的研究热点。机器学习由于其强大的计算能力，近年来已成为众多领域开展数据分析、数据挖掘、文本处理、图像识别等工作的常用方法，本文将针对生态环境损害鉴定评估工作流程中各环节中的难点，探究机器学习模型在各环节中可以起到的关键作用，以期为鉴定评估工作的有序化、系统化和高质量发展提供参考。

2 机器学习介绍

2.1 传统机器学习模型

传统的机器学习模型可根据学习方式的不同分为监督学习和无监督学习[4]，二者的差异及工作方式如图1所示。监督学习是指通过算法建立已知数据中特征（输入）和标签（输出）之间的映射关系，进而对新输入的数据进行预测。根据标签类型的不同，监督学习可以分为回归预测与分类预测，分类预测是指数据的标签被分成不同的类别，而回归预测中数据标签为连续的数值。监督学习常见算法包括决策树[5]、随机森林[6]、XGBoost[7]、支持向量机[8]等。无监督学习是指通过算法挖掘无标签数据之间的内在关系，从中发现数据的规律和模式，无监督学习常见算法包括主成分分析（PCA）[9]、k-means聚类[10]等。

2.2 深度学习与生成式模型

深度学习[11]属于机器学习的分支，主要基于神经网络算法，从已有数据中学习特征和模式，并对新的数据进行分类或预测，与传统的机器学习不同，深度学习可以学习多层抽象特征，因此可以处理更加复杂且高维的数据，也具有更高的预测精度。神经网络由多层神经元组成，每层神经元接受上一层多个神经元的输入，并输出一个结果给下一层的神经元，通过增加神经网络的层级，可以学习更加复杂抽象的数据特征，得到更好的预测效果。目前，深度学习的应用非常广泛，如

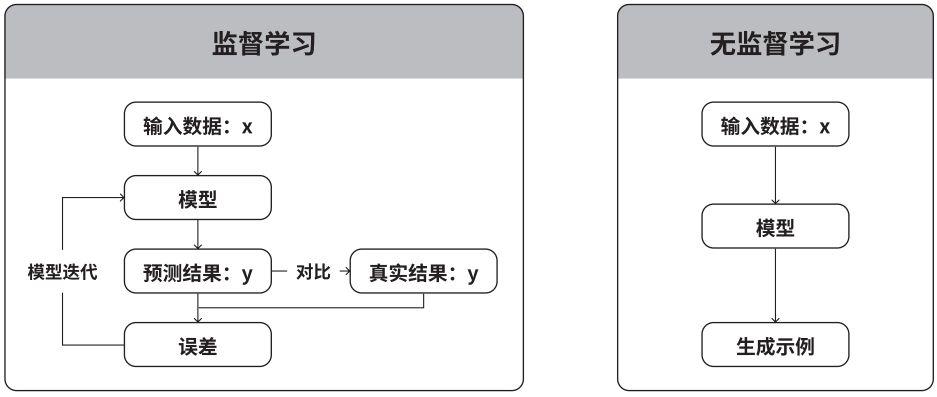

图1 监督学习与无监督学习的工作方式及差异

Figure 1 Working methods and differences between supervised learning and unsupervised learning

计算机视觉[12]、自然语言处理[13]和自动驾驶[14]等。

近年来，随着ChatGPT的出现，生成式人工智能[15]也慢慢被人们所熟知。生成式人工智能是深度学习的分支，其可以根据训练与学习的内容生成新的内容。当收到用户提供的提示词时，生成式人工智能会利用训练的数据生成统计模型预测答案，并产生新的内容来回答问题。生成式人工智能在文本生成（生成写作、聊天机器人、邮件编写等）、图片、视频生成领域都有巨大的应用价值[16]。

3 机器学习在生态环境损害鉴定评估工作流程中的应用前景

《生态环境损害鉴定评估技术指南 总纲和关键环节 第一部分：总纲》（GB/T 39791.1—2020）中介绍了生态环境损害鉴定评估的工作流程，主要包括工作方案制定、损害调查确认、因果关系分析、损害实物量化、损害价值量化、评估报告编制和恢复效果评估7项步骤，流程图如图2所示[17]。下文将结合鉴定评估的主要工作程序，探究机器学习在各项工作流程中的应用前景。表1列举了机器学习模型在生态环境损害鉴定评估各工作流程中可发挥的作用。

3.1 工作方案制定阶段

工作方案制定是生态环境损害鉴定评估工作流程中的第一步，鉴定人需通过收集资料、现场踏勘、座谈走访、文献查阅、遥感影像分析等方式，掌握污染环境或破坏生态行为以及生态环境的基本情况，

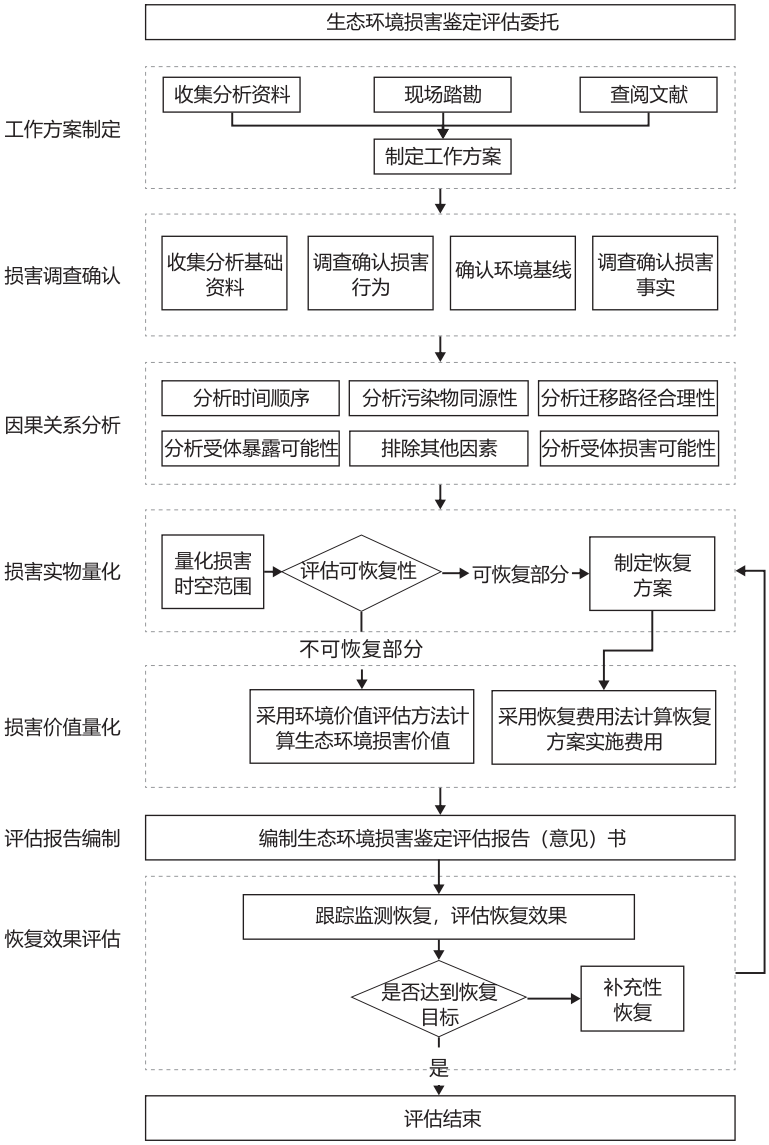

图2 生态环境损害鉴定评估程序图

Figure 2 Diagram of the procedure for identification and assessment of damage to the ecosystem

表1 机器学习在生态环境损害鉴定评估工作中的应用前景

Table 1 Prospects of machine learning in ecological environmental damage identification and assessment work

鉴定评估工作流程 Workflow of appraisal and assessment	机器学习应用方法 Method of machine learning	目标作用 Purpose
资料收集与整理	自然语言处理	提取有效信息
现场勘察	图像识别	水质检测、固废识别、样方调查
损害调查确认	数据挖掘	环境基线确认
因果关系分析	数据挖掘	污染物同源性分析
损害量化	自然语言处理	选取量化方法、预测赔偿金额
鉴定评估报告撰写	自然语言处理	辅助报告撰写、审校与排版
鉴定报告质量评估	自然语言处理	报告质量评估

确定生态环境损害鉴定评估的目的、对象、范围、内容、方法、质量控制和质量保证措施等，编制鉴定评估工作方案。

3.1.1 资料收集与整理

作为鉴定评估工作流程中的首个步骤，资料收集与整理为鉴定评估工作的顺利开展奠定了基础，资料的全面性与准确性对后续损害确认与量化都有显著影响。在实践案例过程中，鉴定人往往会收集大量关于案件的资料信息，其中包括作案人及相关人员口供、检测数据、企业环评报告等，资料内容多且信息类型各异，传统人工梳理资料的方式需要耗费较长时间，且难以从杂乱无章的信息中提取关键有效内容，借助基于人工智能的自然语言处理技术，可以采用同类型的案件资料进行训练，通过分析大量文本内容，学习资料中的规律与特征，帮助鉴定人提升信息筛选效率。同时，基于大语言模型技术[18]，可以通过向模型中导入案件资料内容，实现指定内容的智能问答功能[19]，根据鉴定工作需要，结合鉴定人主动提出的问题，自动对相关信息进行挖掘分析并将答案反馈给鉴定人，帮助鉴定人在杂乱的内容中快速提取有效信息，提升鉴定评估效率。

3.1.2 现场勘察

现场勘察阶段，鉴定人需要通过各种采样及监测手段，判断调查区域的环境要素是否受到不利影响，从而判断调查区域是否存在污染环境或破坏生态的情况。在地表水环境损害鉴定评估工作中，目前水质的检测方式一般采用液相色谱-质谱等仪器，这类传统的水质检测方式精度高、测量效果好，但同时也存在计算量大、数据质量要求高、仪器设备昂贵等局限，使得检测流程复杂、数据获取周期长。近年来，将机器学习模型引入水质检测领域已成为趋势，例如，曹生现等[20]基于纳氏试剂分光光度法，根据水体图像的颜色特征值预测氨氮浓度。另有研究[21]将LM-BP神经网络、支持向量机等模型结合水样的紫外多波段光谱预测COD值。以上案例均是利用机器学习模型预测水质的单一变量，然而实际损害鉴定案例中面临复杂多变的水质情况，单一变量并不能全面反映实际的水质类别。彭子康等[22]首先通过图像处理得到水质颜色特征集，再通过决策树构建图像颜色特征RGB值与水体浑浊度的相关性模型，再基于水体氨氮浓度与浑浊度的关系，构建了水体氨氮浓度与最佳匹配颜色特征G值的关系模型，该方法处理水体图像时能同时检测水体浑浊度及氨氮浓度，开创性地完成多变量水质预测。图像处理的应用在一定程度上为水样检测带来了便利，可以在鉴定评估工作中实现快速原位检测，辅助鉴定人初步判断水质整体情况。

在涉固废类案件中，鉴定工作初期鉴定人需要对固废进行成分检测及含量分析，进而判断固废类型并评估其对环境的污染情况，目前部分研究也已将机器学习模型用于固废识别领域。Brisola等[23]利用经典监督分类模型，结合建筑垃圾的尺寸、形状和纹理特征，对不同来源的建筑垃圾颗粒进行分类，总体正确率达到了73.96%，实现了基于机器学习模型的建筑固废垃圾中细小颗粒的分类。Xia等[24]利用PCA-AdaBoost模型分类预测建筑垃圾中的不同材料，并使用中心均值和K折交叉验证优化模型结果。固体废物中的夹杂物一般通过扫描电子显微镜（SEM）和能谱仪（EDS）进行定性检测，但使用过程时间成本较高且样本具有随机性，梁印等[25]采用图像识别技术，利用Mask-RCNN目标检测算法，实现了对4类夹杂物（AlN、Al_2O_3、MnS和AlN-MnS_4）的准确定位和图像分割。模型对夹杂物的检测识别效果正确率高，对MnS和AlN-MnS_4的识别准确率达到100%，AlN和Al_2O_3夹杂物的识别准确率大于80%。通过图像处理技术识别固废类型，可以有效提高涉固废类鉴定评估的工作效率，帮助鉴定人实现现场勘察过程中的固废种类快速定性，对后续的损害确认及量化起到良好的辅助作用。

在生态环境损害鉴定评估过程中，对涉及生态破坏类的案件，鉴定人需要对调查区域内的植物进行样方调查，了解调查区的植被类型及分布，从而确定优势物种，为后续的损害量化提供依据，同时也需要对周边未受污染的区域开展调查以确认环境基线。在样方调查过程中，鉴定人需要根据自身经验及查阅资

料确定植被类型，当区域内存在的植被种类过多时，样方调查工作耗时耗力，因此，快速确定植被种类可以显著提高样方调查的效率。Ghazi等[26]集成了VGGNet和GoogLeNet，率先实现了在真实情境下的植物识别能力。张露[27]使用AlexNet、Inception-V3、VGG-16预训练模型对植物叶片进行识别，模型在测试集上的预测准确率分别达到了95.31%、93.86%、95.40%。由此可见，基于深度学习神经网络的图像识别技术在植物种类识别领域具有极佳的预测精度以及良好的运用前景，能够显著提高鉴定人的样方调查工作效率。

3.2 损害调查确认阶段

在环境损害鉴定评估工作流程中的损害调查确认阶段，鉴定人调查并对比生态环境及其服务功能现状和基线，确定生态环境损害的事实及其类型，基线的确认方法主要包括参照历史数据、参照对照数据、参照标准基准和开展专项研究，当无法获取历史数据和对照数据且无可用的环境质量标准或基准时，可通过开展专项研究的方式确定环境基线。专项研究过程中，可以利用模型推算的方法预测环境基线，模型推算是通过大量数据构建污染物浓度与生物因子（生物毒性、生物量等）之间的剂量-效应关系模型。早在2011年，Puzyn等[28]使用机器学习模型预测纳米材料的生物效应。Karr等[29]通过构建生物因子与人类活动关系的预测模型有效揭示了环境在未受人为干扰情况下的基本状况。机器学习算法支持处理大体量、多维、高异质性数据[30]，在构建模型的过程中可以考虑多种因素对结果的影响，确保了结果的可靠性与合理性。此外，随着国家环境领域基础数据库的不断完善，机器学习模型将在环境基线确定工作中发挥愈发重要的作用。

3.3 因果关系分析

因果关系分析旨在明确污染环境行为与生态环境损害之间的因果关系，是环境损害鉴定评估工作中至关重要的环节[31]，其主要内容包括判断污染环境行为与生态环境损害发生的时间顺序、判断污染源与环境介质或生物中的污染物是否具有同源性等，其中污染物的同源性分析是因果关系分析中最为重要的部分，须明确污染源与污染区域的污染物是否具有同源性才能确定环境污染责任主体。

目前关于污染物同源性分析主要采用稳定同位素[32-33]、放射性同位素[34]、指纹图谱[35]、多元统计分析[36]等技术方法，目前已有许多研究将机器学习与前述方法相结合，对重金属污染进行溯源分析，如宋志廷等[37]结合随机森林与传统数理统计模型，对天津市武清区土壤重金属污染进行研究，得到Cu、Cr、Cd、Hg、As、Ni、Zn、Pb 8种重金属的污染特征和污染来源。刘斌等[38]采用随机森林模型对晋中盆地土壤中Pb、As、Hg、Cd和Cr 5种重金属污染情况进行建模分析，模型结果表明土壤自然成土过程、工业和农业生产活动、道路交通等是土壤重金属污染的主要来源。除此之外，有研究表明机器学习模型在污染物同源性分析过程中优于传统线性模型，Hu等[39]以长江三角洲作为研究区，对比了随机森林和广义线性模型在土壤重金属污染风险和污染源方面的精度，结果表明前者的模型精度显著大于后者。

3.4 损害量化

在生态环境损害类案件的鉴定评估工作中，环境损害价值量化结果是被告人或当事人最为关心，同时也是鉴定评估工作中需要解决的重要问题之一。环境损害价值量化，即当事人需为环境损害行为所支付的损害赔偿金额[40]。针对这一问题，首先，基于人工智能的自然语言处理技术可以通过分析过往的鉴定评估报告或相关判决文书，结合案件类型、案件基本信息等因素，综合分析不同案件的价值量化评估方法，基于本次案件特征，推荐合适的价值量化方法；根据案件涉及的污染量、污染特征等因素，结合过往案例量化过程与量化结果，对本次案件的价值量化结果，即对环境损害赔偿金额进行预测。由于环境损害类案件涉及的环境要素多、污染物类型复杂多样，不同的案件涉及到的价值量化方法与量化过程均存在一定差异，因此，需要构建环境损害类案例数据库，结合大量案件信息进行分析，方可保证预测结果的准确性与可靠性。

目前，与损害赔偿金额预测类似的研究方向——刑期预测已存在部分应用实例[41]。近年来部分法院与科技公司建立合作，研发人工智能量刑工具，例如，贵州高级人民法院开发的法镜大数据系统，其通过对犯罪事实进行精确“画像”，构建“要素-证据-量刑”关联模型，通过相似案件进行对比，为法官提供量刑建议。海南高级人民法院利用量刑规范化智能辅助系统，对导入的起诉书、辩护词等诉讼材料进行智能分析，挖掘案件基本信息并根据历史案件信息推荐量刑[42]。据此，环境损害价值量化与刑期预测具有相同的工作原理，其同样是通过构建人工智能模型对过往大量案例进行挖掘分析，得出预测结果。因此，随着

环境损害类案例数量的不断增多，构建相关案例数据库并对过往案例进行挖掘分析，通过自然语言处理技术推荐损害价值量化方法并预测损害赔偿金额，能够为鉴定人提供有效参考，避免出现同类案件量化结果差异巨大的情况，确保了鉴定评估工作的公正性与合理性。

3.5 评估报告编制

3.5.1 辅助鉴定评估报告撰写

鉴定评估报告的撰写要求鉴定人结合上述所有流程中的资料、信息和数据进行分析，输出鉴定评估结果，对于不同类型的环境损害类案件，一般采用固定的报告模板进行撰写，整体的撰写思路也与鉴定评估的工作流程相同。目前国内环境损害类案件的数量呈现逐年递增的趋势[2]，大量案件的发生导致鉴定人的工作量激增，鉴定人既需要保证鉴定评估报告的有效性、科学性与合理性，也需要在委托人规定的时间内完成报告内容撰写，因此，随着案件数量的增多，鉴定人亟需一款文本智能生成工具协助完成报告撰写工作[43]。

随着深度学习的逐步发展，目前已有许多研究人员利用深度学习技术开展机器写作工作，例如自动生成文章摘要，在20世纪90年代，研究人员提出生成式文本摘要方法（Abstractive Text Summarization，ATS），该类方法利用神经网络模型模拟人类书写习惯，通过训练生成文本摘要[44]。随后，Nallapati等[45]提出一种Attention的条件随机场模型，通过模仿人类的决策方式，能够生成更加书面化的文本摘要。Lopyrev[46]提出一种基于包含长短期记忆模型（Long Short-Term Memory，LSTM）单元的循环神经网络（Recurrent Neural Networks，RNN）结构和Attention机制的ATS方法，该方法首先计算文本语句相关性和冗余度，其次评估子句单元的相关性，然后将两者作为RNN结构和Attention机制的输入，既提升了文本摘要的生成质量，同时也缩短了模型的训练时间。在环境损害鉴定评估报告撰写工作中，损害调查确认和鉴定评估结论模块均可由报告前文的具体资料和信息凝练总结得到，因此，基于深度学习的ATS技术，可以协助鉴定人快速完成鉴定报告部分内容的撰写。

与此同时，对于部分大型或复杂案件，其涉及的鉴定评估流程复杂，鉴定评估报告篇幅较长，后期的报告加工与审校需要的人力和时间成本较高，大语言模型的发展极大提高了计算机对文本的语义分析和上下文联系能力，通过构建环境损害鉴定评估领域的知识库，联网的大语言模型可以智能判断各类知识点的正误，其可在语言书写逻辑及语法矫正等方面为鉴定报告撰写人提供提醒和参考[47]。此外，大语言模型还能识别报告中的标点符号、公式、插图表格、大小写、角标等格式规范问题[48]，显著提高报告排版与审校的工作效率。

3.5.2 鉴定报告质量审核

目前生态环境损害鉴定评估报告的质量审核主要依靠专家咨询方式，易受评阅专家自身经验及主观判断影响，导致不同专家出示的专家意见不统一，即使通过标准化的分级权值对报告进行评价，分数可能也会有较大差异。并且，随着生态环境损害类案件的数量不断增多，专家难以对所有鉴定报告进行通篇的仔细阅读，从而导致报告质量审核的准确性难以控制。目前，国内外学者对文章内容的客观准确评定与分级也在进行深入研究，基于自然语言处理的内容编码和理解与评估算法对文章质量智能评估方法也提供了必要的技术基础。文章质量智能评估的任务是根据文章中与质量相关的特征，经过计算对文字进行评分，可以自动判断文章质量，进而减轻鉴定评估专家对报告质量审核的负担[49]。在鉴定评估领域，首先需要构建鉴定评估报告的数据库，利用模型提取报告文本内容，学习文本特征如报告框架、报告结构等，进而对新的鉴定报告进行质量评估。可以预见的是，随着自然语言处理算法的不断精进、鉴定评估报告相关数据库及质量评估指标体系的不断完善，模型会对报告的质量进行更准确、更专业的评估。

4 机器学习在生态环境损害鉴定评估工作中的应用局限性

机器学习在生态环境损害鉴定评估中具有广泛的应用潜力，但也存在一定的局限性。首先，机器学习模型对数据依赖较强，需要大量的优质数据进行训练和优化。然而，在生态环境损害鉴定评估领域，数据的获取可能受到限制，会出现数据不足、数据质量参差不齐等问题，这可能会对机器学习模型的准确性和可靠性造成一定影响。充足且高质量的数据资源是机器学习模型训练的基础，因此，在生态环境损害鉴定评估工作中，只有当鉴定人拥有充足且高质量的历史数据或对照数据时，模型结果才能对鉴定评估结果起到有效、可靠的支持。其次，机器学习模型的泛化能力也是需要考虑的局限之一，机器学习模型通常基于训练数据进行拟合，但在面对新数据或不同情境

下的数据时，其表现可能不够理想。在生态环境损害鉴定评估中，由于环境因素的复杂性和变化性，模型的泛化能力可能会受到限制，无法适应各种类型的生态环境损害案件。为克服上述问题，首先应构建大型的生态环境损害鉴定评估领域数据库，例如水质图像数据库、固废图像数据库、污染物溯源数据库、污染物特征指纹数据库、生态环境损害鉴定案例数据库等，通过模型对大量数据的训练、预测及外部验证，不断提高模型预测精度，确保模型预测结果的准确性和有效性。

此外，由于深度学习神经网络的内部结构异常复杂，其运行机制类似一个“黑箱”[50]，故其也被称作黑箱模型，这导致模型预测的方式难以用人类可以理解的语言去描述，模型输出的结果也难以被解释，而生态环境损害鉴定评估要求鉴定人需要按照规定的程序和方法开展工作，对鉴定评估结果的准确性及严谨性要求极为严格，因此，由黑箱模型得出结论，其可信度始终会受到内部及外界的挑战。尤其在环境损害司法鉴定过程中，鉴定人所出具的鉴定报告需要在法庭中作为判案证据，机器学习模型得到的结果在证据可采性方面较易受到法官及当事人的挑战与质疑。

为了克服这些局限性，我们可以采取一系列措施。首先，加强生态环境损害领域的数据收集和预处理工作，确保数据的质量和可比性。其次，结合领域知识进行特征选择，以捕捉评估中的重要因素。此外，可以运用可解释性更强的机器学习模型，以提高预测结果的可解释性和可信度。最终，鉴定人应该将机器学习方法与专家知识和经验相结合，综合评估生态环境的损害情况，以实现更准确、可靠的鉴定评估结果。

5 结论

本文探究了机器学习在生态环境损害鉴定评估中的应用前景，结合鉴定评估的主要工作流程：工作方案制定、损害调查确认、因果关系分析、损害实物量化、损害价值量化和评估报告编制，介绍了机器学习在各流程中可以发挥的关键作用，旨在提高鉴定评估工作效率，推动鉴定评估工作有序化、系统化及高质量发展。此外，虽然机器学习在数据挖掘、图像识别、自然语言处理应用中具备较好的模型预测效果，也在众多领域取得了不错的成果，但是其在生态环境损害鉴定评估中的应用也存在一定的局限性，如模型对数据的依赖性较强、泛化能力不足以及模型结果的可解释性不足等。未来，随着计算机技术的不断进步和创新以及生态环境损害鉴定评估领域的数据库逐步建立，机器学习在生态环境损害鉴定评估中的应用将得到广泛发展。

参考文献：

[1] 於方, 张衍燊, 齐霁, 等. 环境损害鉴定评估关键技术问题探讨[J]. 中国司法鉴定, 2016(1):18-25. YU F, ZHANG Y S, QI J, et al. Key technical issues in environmental damage appraisal and assessment[J]. *Chinese Journal of Forensic Sciences*, 2016(1):18-25.

[2] 於方, 张衍燊, 徐伟攀.《生态环境损害鉴定评估技术指南 总纲》解读[J]. 环境保护, 2016, 44(20):9-11. YU F, ZHANG Y S, XU W P. Interpretation of *the General Principles of Technical Guide for Ecological and Environmental Damage Assessment*[J]. *Environmental Protection*, 2016, 44(20):9-11.

[3] 马苗. 生态环境损害惩罚性赔偿数额的确定[J]. 中华环境, 2021(6):73-76. MA M. Determination of the amount of punitive damages for ecological environment damage[J]. *China Environment*, 2021(6):73-76.

[4] SIZOCHENKO N, SYZOCHENKO M, FJODOROVA N, et al. Evaluating genotoxicity of metal oxide nanoparticles: application of advanced supervised and unsupervised machine learning techniques[J]. *Ecotoxicology and Environmental Safety*, 2019, 185:109733.

[5] LOH W. Classification and regression trees[J]. *WIREs Data Mining and Knowledge Discovery*, 2011, 1(1):14-23.

[6] Random Forest: a classification and regression tool for compound classification and QSAR modeling | journal of chemical information and modeling[EB/OL].[2023-09-18]. https://pubs.acs.org/doi/10.1021/ci034160g.

[7] XGBoost | proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining[C/OL]. /[2023-09-18]. https://dl.acm.org/doi/10.1145/2939672.2939785.

[8] BURDEN F R, WINKLER D A. Relevance vector machines: sparse classification methods for QSAR[J]. *Journal of Chemical Information and Modeling*, 2015, 55(8):1529-1534.

[9] MULLER K R, MIKA S, RATSCH G, et al. An introduction to kernel-based learning algorithms[J]. *IEEE Transactions on Neural Networks*, 2001, 12(2):181-201.

[10] JAIN A K. Data clustering: 50 years beyond K-means[J]. *Pattern Recognition Letters*, 2010, 31(8):651-666.

[11] Deep learning | Nature[EB/OL]. [2023-09-18]. https://www.nature.com/articles/nature14539.

[12] KER J, WANG L P, RAO J, et al. Deep learning applications in medical image analysis[J]. *IEEE Access*, 2018, 6:9375-9389

[13] 赵京胜, 宋梦雪, 高祥. 自然语言处理发展及应用综述[J]. 信息技术与信息化, 2019(7):142-145. ZHAO J S, SONG M X, GAO X. An overview of natural language processing development and applications[J]. *Information Technology and Informatization*, 2019(7):142-145.

[14] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. *Journal of Field Robot-*

ics, 2020, 37(3):362-386.

[15] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.

[16] SALLAM M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns[J]. *Healthcare, Multidisciplinary Digital Publishing Institute*, 2023, 11(6):887.

[17] 生态环境部. 生态环境损害鉴定评估技术指南 总纲和关键环节 第1部分:总纲GB/T 39791. 1—2020[S]. 生态环境部, 2022. Ministry of Ecology and Environment. Technical guidelines for identification and assessment of environmental damage: general principles and key components —Part 1: general principles GB/T 39791.1—2020 [S]. Ministry of Ecology and Environment, 2022.

[18] 刘明, 吴忠明, 廖剑, 等. 大语言模型的教育应用:原理、现状与挑战:从轻量级BERT到对话式ChatGPT[J]. 现代教育技术, 2023, 33(8):19-28. LIU M, WU Z M, LIAO J, et al. Educational applications of large language models: principles, status and challenges-from light-weight BERT to conversational ChatGPT[J]. *Modern Educational Technology*, 2023, 33(8):19-28.

[19] 张鹤译, 王鑫, 韩立帆, 等. 大语言模型融合知识图谱的问答系统研究[J]. 计算机科学与探索, 2023, 17(10):2371-2388. ZHANG H Y, WANG X, HAN L F, et al. Research on question answering system on the joint of knowledge graph and large language models[J]. *Journal of Frontiers of Computer Science and Technology*, 2023, 17(10):2371-2388.

[20] 曹生现, 王延红, 郑丽婷, 等. 基于图像处理的水质中氨氮含量的检测研究[J]. 化工自动化及仪表, 2014, 41(8):910-914. CAO S X, WANG Y H, ZHENG L T, et al. Research on ammonia nitrogen content detection based on image processing technology[J]. *Control and Instruments in Chemical Industry*, 2014, 41(8):910-914.

[21] 宓云軿, 王晓萍, 金鑫. 基于机器学习的水质COD预测方法[J]. 浙江大学学报(工学版), 2008(5):790-793. MI Y P, WANG X P, JIN X. Water COD prediction based on machine learning[J]. *Journal of Zhejiang University(Engineering Science)*, 2008(5):790-793.

[22] 彭子康, 张芹, 陈世航, 等. 基于机器学习和图像处理的水质检测方法[J]. 自动化应用, 2023, 64(10):188-191. PENG Z K, ZHANG Q, CHEN S H, et al. Water quality detection methods based on machine learning and image processing[J]. *Automation Application*, 2023, 64(10):188-191.

[23] BRISÖLA D F, CUNHA B M, GOMES O. Automatic classification of particles from construction and demolition waste through digital image analysis[C]. 65th AMB International Congress, 2010, 4: 3046-3052.

[24] XIA H, BAKKER M C M. Reliable classification of moving waste materials with LIBS in concrete recycling[J]. *Talanta*, 2014, 120: 239-247.

[25] 梁印, 朱航宇, 罗林根, 等. 基于深度学习钢中非金属夹杂物图像识别[J/OL]. 钢铁, doi: 10.132281j.boyuan.issn0449-749X.20230165. LIANG Y, ZHU H Y, LUO L G, et al. Image recognition of non-metallic inclusions based on deep learning[J/OL]. *Steel*, doi: 10.132281j.boyuan.issn0449-749X.20230165.

[26] GHAZI M M, YANIKOGLU B A, APTOULA E. Open-set plant identification using an ensemble of deep convolutional neural networks[C]. Portugal: Working Notes of Conference and Labs of the Evaluation Forum, 2016:18-524.

[27] 张露. 基于深度学习的植物叶片图像识别方法研究[D]. 北京:北京林业大学, 2019. ZHANG L. The study of plant leaf image recognition based on deep learning method[D]. Beijing: Beijing Forestry University, 2019.

[28] PUZYN T, RASULEV B, GAJEWICZ A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles[J]. *Nature Nanotechnology*, 2011, 6(3):175-178.

[29] REES W E. Restoring life in running waters: better biological monitoring.[J]. *Ecosystem Health, Malden*, 2000, 6(2):166-168.

[30] 武子豪, 张司雨, 董仕鹏, 等. 机器学习在纳米材料风险评估中的应用[J]. 生态毒理学报, 2022, 17(5):139-151. WU Z H, ZHANG S Y, DONG S P, et al. Overview of application of machine learning in field of nanomaterials risk assessment[J]. *Asian Journal of Ecotoxicology*, 2022, 17(5):139-151.

[31] 陈伟. 生态环境损害责任因果关系及其证明[J]. 南京工业大学学报(社会科学版), 2023, 22(4):17-28. CHEN W. The causality of ecological environment damage liability and its proof[J]. *Journal of Nanjing University of Technology(Social Science Edition)*, 2023, 22(4):17-28.

[32] 梁增强, 毛安琪, 杨菁. 生态环境损害鉴定评估技术难点探讨[J]. 环境与发展, 2019, 31(12):241-242. LIANG Z Q, MAO A Q, YANG J. Analysis of technical difficulties of assessment on eco-environmental damage[J]. *Environment and Development*, 2019, 31(12): 241-242.

[33] 王万洁, 侯兴旺, 刘稷燕, 等. 传统稳定同位素技术在环境科学领域的应用及研究进展[J]. 环境化学, 2021, 40(12):3640-3650. WANG W J, HOU X W, LIU J Y, et al. Application and research progress of traditional stable isotope technology in environmental science [J]. *Environmental Chemistry*, 2021, 40(12):3640-3650.

[34] 欧阳洁, 杨国胜, 马玲玲, 等. 大气污染物中人工放射性铯-钚-铀同位素示踪技术的发展与应用[J]. 化学进展, 2017, 29(12):1446-1461. OUYANG J, YANG G S, MA L L, et al. Development and application of fingerprints of radioactive cesium-plutonium-uranium isotopes as tracers in air pollution[J]. *Progress in Chemistry*, 2017, 29(12):1446-1461.

[35] 余明星, 苏海, 张凯, 等. 基于有机物指纹图谱技术的长江南京段水中有机磷溯源研究[J]. 环境监控与预警, 2022, 14(5):100-106. YU M X, SU H, ZHANG K, et al. Research on tracing organophosphorus in Nanjing section of Yangtze river based on fingerprint technology of organic compounds[J]. *Environmental Monitoring and Forewarning*, 2022, 14(5):100-106.

[36] 尚婷婷, 张亚群, 周静, 等. 多元统计分析在农田土壤重金属污染源解析中的应用[J]. 环境生态学, 2022, 4(4):93-97. SHANG T T, ZHANG Y Q, ZHOU J, et al. Application of multivariate statistical analysis in the analysis of heavy metal pollution sources in farmland soil[J]. *Environmental Ecology*, 2022, 4(4):93-97.

[37] 宋志廷, 赵玉杰, 周其文, 等. 基于地质统计及随机模拟技术的天津武清区土壤重金属源解析[J]. 环境科学, 2016, 37(7):2756-2762. SONG Z T, ZHAO Y J, ZHOU Q W, et al. Applications of geostatistical analyses and stochastic models to identify sources of soil heavy metals in Wuqing district, Tianjin, China[J]. *Environmental Science*, 2016, 37(7):2756-2762.

[38] 刘斌, 郭星, 朱宇恩. 基于随机森林模型的土壤重金属源解析:以晋中盆地为例[J]. 干旱区资源与环境, 2019, 33(1):106-111. LIU B, GUO X, ZHU Y E. Analysis of soil heavy metal sources in Jinzhong basin based on random forest model[J]. *Journal of Arid Land Resources and Environment*, 2019, 33(1):106-111.

[39] HU B, XUE J, ZHOU Y, et al. Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning[J]. *Environmental Pollution*, 2020, 262: 114308.

[40] 张辉. 生态环境损害惩罚性赔偿金数额的计算基数:困境与出路[J]. 法学论坛, 2023, 38(5):110-123. ZHANG H. On the calculation base of punitive damages for ecological environment damage:predicament and outlet[J]. *Legal Forum*, 2023, 38(5):110-123.

[41] 戴非凡, 司卫云, 倪进平. 基于机器学习的司法数据分析及建模研究:以"故意伤害罪"为例[J]. 智能计算机与应用, 2020, 10(1):89-94. DAI F F, SI W Y, NI J P. Analysis and modeling of judicial data based on machine learning:take the crime of intentional injury as an example[J]. *Intelligent Computer and Applications*, 2020, 10(1):89-94.

[42] 刘雁鹏. 中国司法大数据应用的主要成就与理论逻辑[J]. 齐鲁学刊, 2023(4):96-109. LIU Y P. The main achievements and theoretical logic of judicial big data in China[J]. *QiLu Journal*, 2023(4): 96-109.

[43] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理:挑战、机遇与发展[J]. 中国科学:信息科学, 2023, 53(9):1645-1687. CHE W X, DOU Z C, FENG Y S, et al. Towards a comprehensive understanding of the impact of large language models on natural language processing:challenges, opportunities and future directions[J]. *Scientia Sinica Informationis*, 2023, 53(9):1645-1687.

[44] Improving sentence compression by learning to predict gaze - ACL Anthology[EB / OL]. [2023-09-18]. https://aclanthology. org / N16-1179/.

[45] NALLAPATI R, ZHOU B, SANTOS C N DOS, et al. Abstractive text summarization using sequence-to-sequence RNNs and beyond[DB]. arXiv, 2016.

[46] LOPYREV K. Generating news headlines with recurrent neural networks[DB]. arXiv, 2015.

[47] 黄杰. 自然语言处理在文本审核中的应用[J]. 网络安全技术与应用, 2021(3):27-29. HUANG J. Application of natural language processing in text review[J]. *Network Security Technology and Application*, 2021(3):27-29.

[48] 王峻峰. 人工智能生成内容(AIGC)及其在图书出版中的应用探讨[J]. 传播与版权, 2023(10):48-51. WANG J F. Exploration of artificial intelligence generated content(AIGC) and its application in book publishing[J]. *CHUANBO YU BANQUAN*, 2023(10):48-51.

[49] 黄鹏程. 基于深度学习的文章质量评估方法研究[D]. 重庆:西南科技大学, 2023. HUANG P C. Research on article quality evaluation method based on deep learning[D]. Chongqing:Southwest University of Science and Technology, 2023.

[50] RIBEIRO M T, SINGH S, GUESTRIN C. *Why Should I Trust You?*: explaining the predictions of any classifier[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, California, USA:ACM, 2016:1135-1144.