

人工智能辅助食品安全主动防控研究进展

盛利娜¹, 纪剑¹, 宋晓宁², 吴小俊², 孙秀兰^{1*}

¹ 江南大学食品学院 江苏无锡 214122

² 江南大学人工智能与计算机学院 江苏无锡 214122)

摘要 食品安全事关全球公共健康。面对新业态、新资源食品中的新发、突然危害物,食源性致病菌和真菌毒素污染以及化学残留等传统危害物的多重挑战,人工智能(AI)技术的飞速发展提供了新的解决思路。本文系统总结 AI 技术在食品安全主动防控领域的应用,包括风险预警、毒性预测、快速检测和高效防控。阐述 AI 技术综合气象统计方法、机理模型和机器学习算法,实现对食品中致病菌、真菌毒素、农残和重金属等风险因子的早期预警。针对新发、突发危害物,将 AI 与传统毒理学模型相融合,结合迁移学习,实现对危害物的毒性预测和风险评估。AI 除了广泛应用于食品安全和食品品质的快速检测外,还能够辅助抗体和适配体等传统和新型识别元件的高效设计与筛选。在主动防控领域,AI 与生物信息学、分子生物学和合成生物学相结合,可对抗菌肽、降解酶和噬菌体等进行高效预测与筛选并揭示其防控机制。AI 技术在食品安全领域的应用仍面临数据共享不足,标准化、多模态数据处理等挑战。随着技术的不断完善和数据共享机制的改进,AI 将在保障全球食品安全,应对复杂多变的食品安全问题上发挥更为重要的作用。

关键词 人工智能; 食品安全; 风险预警; 毒性预测; 快速检测; 高效防控

文章编号 1009-7848(2024)10-0014-14 **DOI:** 10.16429/j.1009-7848.2024.10.002

食品安全作为公共健康的重要组成部分,一直受到全球的广泛关注,保障食品质量与安全已成为各国政府和科研机构的共同目标。造成食品损失及安全问题的原因主要有食源性致病菌、真菌毒素、腐败变质、化学残留等。食源性疾病通常具有传染性或毒性,由细菌、病毒、寄生虫或化学物质,经受污染的食物或水进入人体后所致。2018 年,世界银行关于食源性疾病经济负担的报告指出,低收入和中等收入国家与食源性疾病相关的总生产力损失约为每年 952 亿美元,每年用来治疗食源性疾病的费用约为 150 亿美元。婴幼儿、孕妇、老人以及患有原发疾病的人群尤其脆弱。每年有 2.2 亿儿童患腹泻病,9.6 万人死亡^[1]。2022 年,所有生产的食品中有 13% 被损失,另外 19% 的食品在零售、餐饮服务和家庭层面被浪费^[2]。全球每年收获的农作物中约有 25% 受到真菌毒素的污染,其中约 2% 因受到严重污染而失去营养和经济

价值,在农业和工业领域造成数百亿美元的损失^[3]。据《2022 年中国食物与营养发展报告》显示,我国食物总体损耗率为 14.7%,高于世界平均水平,浪费量达 3 亿 t。我国储粮的平均损失率为 8%,每年因黄曲霉毒素污染而造成的粮食损失超过 0.2 亿 t^[4],解决食品安全及损失问题迫在眉睫。

近年来,人工智能(Artificial intelligence, AI)技术飞速发展,利用 AI 解决食品安全问题成为一个新的热点。AI 是指让机器具有类似人类智能的技术,其核心目标是让计算机能够执行人类的任务,从而自动、智能、准确地处理复杂任务,提高工作效率。机器学习(Machine learning, ML)是 AI 的重要组成部分,应用不同算法对大量数据进行分析计算,拟合数据潜在分布并预测未来数据或趋势。监督学习算法是最常见的机器学习算法,包括但不限于线性回归、逻辑回归、决策树、支持向量机、K-近邻和随机森林算法。AI 技术已被用于食品生产加工的全链条,如食品分级、分类和质量评价等^[5]。在食品安全领域,AI 在风险识别和智能监管等方面的应用已有相关综述^[6-10],然而,其在食品危害物主动防控领域的应用尚无全面的综述报道。本文系统总结 AI 在食品中新发、突发和传统危害物的风险预警、毒性预测、快速检测和高效防

收稿日期: 2024-10-29

基金项目: 国家重点研发计划项目(2023YFF1105102, 2023YFF1105105); 国家杰出青年科学基金项目(32125031)

第一作者: 盛利娜,女,博士,副教授

通信作者: 孙秀兰 E-mail: sxlzzz@jiangnan.edu.cn

控领域的应用,以期为 AI 辅助食品安全主动防控提供理论支撑。

1 风险因子早期预警

气候变化对粮食安全构成了巨大威胁。据估计,1850 年至 1900 年间,全球平均地表温度上升了 0.87 °C,并且在未来二三十年内可能再升高 0.5 °C^[11]。全球变暖及其引发的干旱、洪涝、风暴、高温和冷害等极端气候,造成全球每年损失 1 430 亿美元^[12]。高温和干旱会降低作物的产量和质量,增加作物病虫害的风险,暴雨和洪水也会导致土壤侵蚀和作物损毁,进一步威胁粮食供应^[13]。气候变化还会增加粮食被真菌毒素污染的风险^[14-15]。温度升高使得嗜热和耐热真菌得以迁移、引入并逐渐增多,产毒真菌,如黄曲霉和禾谷镰刀将很快适应新的环境,从而成为侵染能力更强的病原体^[14]。此外,气候变化还会削弱粮食作物的防御能力,从而增加真菌感染和产毒的可能性^[16]。目前国内对真菌毒素的预测研究相对较少,国外在这一领域的研究起步较早,已建立一些关于作物真菌毒素污染的预测方法与模型,主要包括气象统计方法、机理模型和机器学习模型^[17]。例如,Liu 等^[18]基于 2001 至 2003 年间荷兰 625 份小麦数据,构建了预测冬小麦中呕吐毒素含量的回归模型、机理模型和贝叶斯网络模型,结果显示贝叶斯网络模型的预测精度最高,可达 84.1%。然而,基于机器学习的样本模型需要大量样本数据支持,具有一定的局限性,其或可与遥感技术相结合,对大区域尺度种植的粮食作物进行真菌毒素预测^[17]。

AI 还被用于食品生产加工过程中食源性致病菌污染风险预测。目前研究人员已开发出众多监督学习模型,用于预测农业用水中的食源性致病菌的存在、污染水平和相关的风险因素^[19-21]。例如,Weller 等^[19]比较了支持向量机、贝叶斯模型、回归模型、随机森林等 20 种机器学习算法对农业用水中沙门氏菌和致病性大肠杆菌的预测效果,结果显示:集成学习构建的模型在预测效果上优于更易解释的学习器构建的模型。Ndraha 等^[21]使用 XGBoost (eXtreme gradient boosting) 算法预测牡蛎养殖场中环境因素对副溶血弧菌丰度的影响。该算法能够有效预测牡蛎和海水中的副溶血弧菌

浓度,而在沉积物中的预测效果不佳;海面温度是影响副溶血弧菌浓度的主要因素,风速增加会减少牡蛎和海水中该细菌的数量,而海水的 pH 值也会影响细菌数量,盐度仅对牡蛎和沉积物中的细菌浓度有影响。

食源性致病菌的耐药性也对食品安全构成严重威胁。耐药致病菌可通过食物链传播给人类,导致感染治疗难度增加,治疗成本上升,严重威胁公共健康。研究表明,水产养殖中多抗生素抗性指标 (Multi-antibiotic resistance index, MAR) 与温度呈正相关,即随着气温的升高,耐药性水平也相应增加,耐药菌感染的水产死亡率更高^[22]。如图 1 所示,Nguyen 等^[23]研究了 AI 结合基质辅助激光解析电离飞行时间质谱 (MALDI-TOF-MS) 预测铜绿假单胞菌耐药性的方法,特别是对于 β -内酰胺类/ β -内酰胺酶抑制剂药物,模型表现出较高的准确性【接受者操作特征曲线下面积 (Area under the receiver operating characteristic, AUROC) 分别为 0.869 和 0.856】和特异性 (分别为 0.925 和 0.897)。研究还引入动态分箱技术,在减少特征数量的同时,保持了预测性能,并通过迁移学习进一步提高 8 种抗菌药物的 AUROC 值,其中,美罗培南的 AUROC 从 0.801 提升至 0.844。

除了生物危害物,AI 也被用于食品中化学危害物的预测。Dong 等^[24]使用 Pyraformer 神经网络模型预测我国小麦中的重金属污染。模型对各省份风险评估指标的预测误差 (RMSE 和 MAE) 均保持在 1 以下,对于最低风险级别的样本,模型的精度和召回率均超过 90%,分别比其它模型高出 25.38%~4.15% 和 18.42%~5.26%。Jiang 等^[25]将 AlexNet 卷积神经网络、支持向量机和 K-近邻 3 种机器学习算法用于检测苹果上的农药残留。其中,基于 AlexNet 的卷积神经网络模型表现最佳,对测试集的平均识别率达到 99.09%,单张图像的检测时间为 0.0846 s,显著优于传统方法,实现了高效、低成本的非接触式农残检测。

2 新发、突发危害物毒性预测

新发、突发危害物的毒性评价和风险评估,是食品安全研究的重要组成部分。传统的毒理学研究方法包括动物实验和体外试验,这些方法耗时、

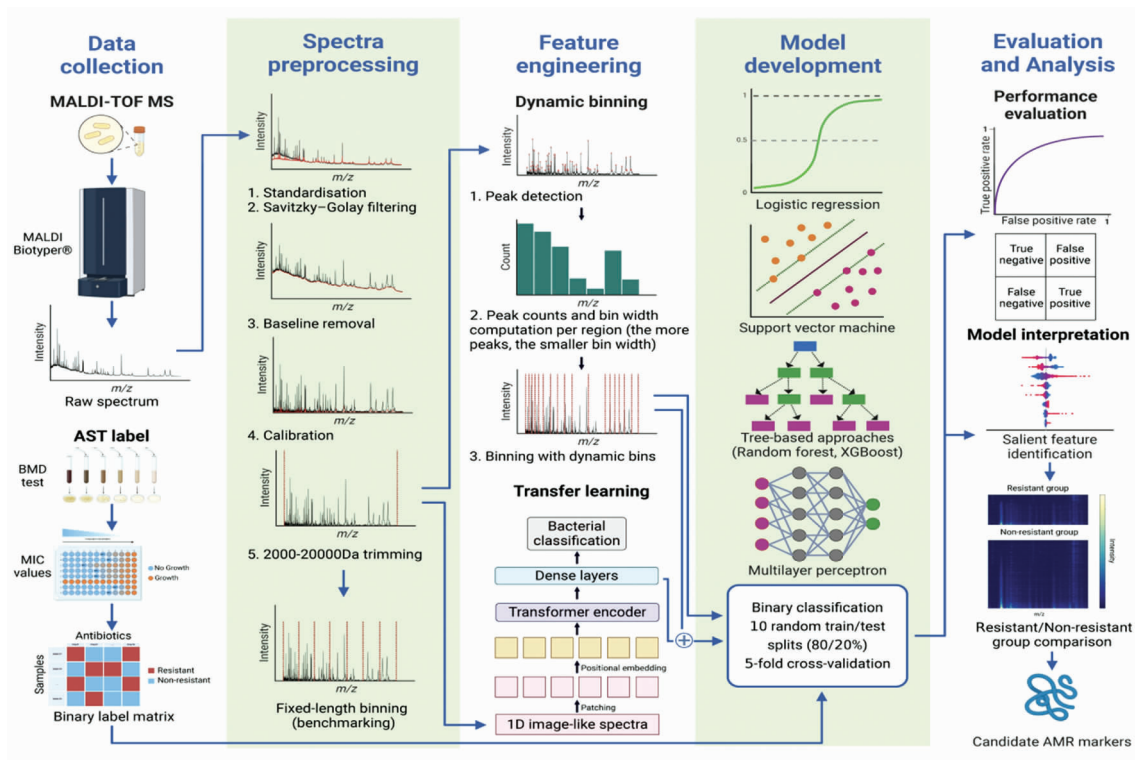


图1 AI结合MALDI-TOF-MS预测铜绿假单胞菌耐药性^[23]

Fig.1 Predicting *Pseudomonas aeruginosa* drug resistance using artificial intelligence and MALDI-TOF-MS^[23]

费力且涉及伦理问题。随着动物实验“3R”原则【即替代(Replacement)、减少(Reduction)和优化(Refinement)】的提出,毒理学预测模型,如生理药代动力学(Physiologically based pharmacokinetic, PBPK)模型、定量构效关系(Quantitative structure-activity relationship, QSAR)模型、有害结局路径(Adverse outcome pathway, AOP)分析和毒物基因组学等被广泛用于危害物的毒性预测^[26]。AI与传统毒理学预测模型相结合,可进一步扩大分析的数据量,从而提高预测效率和准确性。例如,Sharma等^[27]使用多任务深度神经网络结合摩根指纹和预训练的SMILES嵌入,预测了化合物的体外、体内和临床毒性,AUROC值高达0.99以上,且多任务较单任务模型可以更有效地利用跨平台信息,从而提高预测性能(图2)。

QSAR模型常用于研究化合物的结构参数与其生物活性之间的关系。传统QSAR模型多通过试验测定数据进行编码以捕捉其分子特征,并选择合适的数学算法来训练模型,以根据分子描述符来预测化学物质的性质,最后通过验证数据集

评估模型性能。AI结合QSAR模型可以更高效地处理大型数据集,优化特征选择和模型性能,提升化学物质毒性预测的准确性^[28]。AI+QSAR不仅可以判断某化合物是否有毒或无毒,还可以预测剂量-反应关系等^[26]。在食品安全领域,AI+QSAR主要被用于危害物识别、食品添加剂和天然化合物的生物活性评估等^[29]。比如,全氟和多氟烷基物质是一类包含5000多种化合物的大家族,广泛应用于食品工业,可通过食品包装迁移至食物中,威胁人类健康^[30]。Cheng等^[31]构建了一个AI驱动的QASR模型,成功区分3486种全氟和多氟烷基物质的生物活性。作者首先构建了包含1012种全氟和多氟烷基物质的数据库,然后,评估了随机森林、多任务神经网络等传统模型和图卷积网络等高级的基于图的模型等5种机器学习模型在不同生物测定中的表现。结果显示:多任务神经网络和图卷积网络模型预测效果最好,每个生物测定的AUROC均值大于0.9。针对传统QSAR模型无法预测食品化学污染物分子结构与神经毒性的问题,周悦等^[32]利用随机森林、人工神经网络和支持

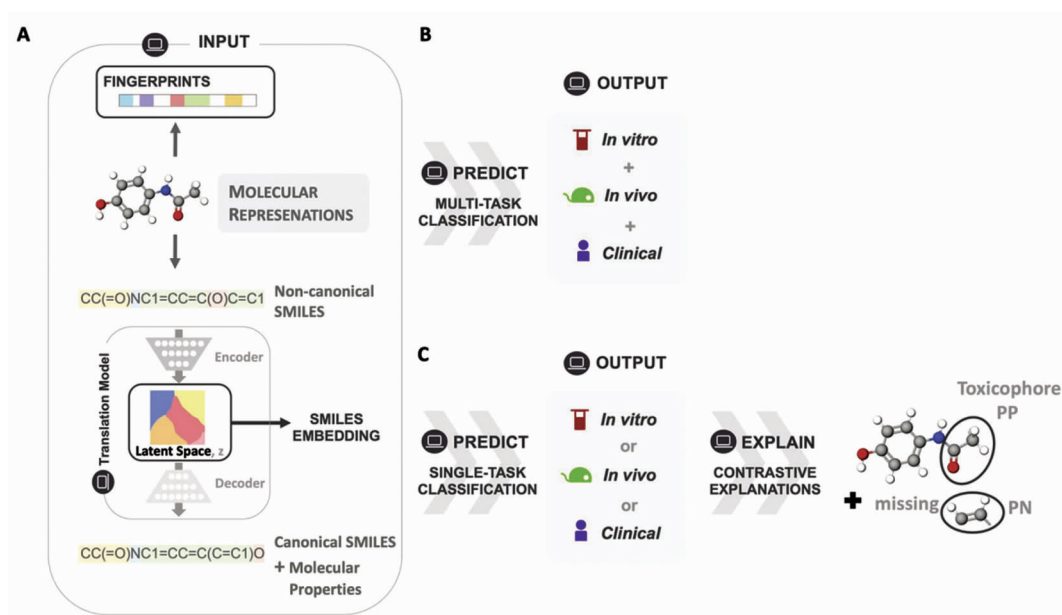


图 2 基于摩根指纹和 SMILES 嵌入的多任务深度学习毒性预测模型^[27]

Fig.2 A multi-task deep learning toxicity prediction model based on Morgan fingerprints and pre-trained SMILES embeddings^[27]

向量机等机器学习算法，构建了预测食品中化学污染物神经毒性的分类模型。首先通过查阅文献建立了包含 57 种影响神经分化成熟,影响神经元迁移/空间定向等各类神经毒性机制化合物和 50 种无神经毒性化合物的化学数据库,经分析比较发现随机森林模型的预测性能最好,其十折交叉验证准确率为 70.24%,训练集和测试集预测准确率分别为 95.51%和 83.33%,AUROC 值分别为 0.99 和 0.85。模型显示,化合物的质量加权 Burden 矩阵最大特征值是预测神经毒性的重要因素。

PBPK 模型则主要用于描述化学危害物在体内的吸收、分布、代谢和排泄过程,广泛应用于剂量-反应分析、暴露评估、体外到体内外推以及物种间外推等毒理学研究领域^[26]。传统 PBPK 模型的构建依赖于大量的试验数据,如组织与血浆分配系数、代谢率等,构建过程时间长、价格高且存在伦理问题。AI 技术可以基于化合物的物理化学性质来预测 PBPK 模型的相关参数,从而大大降低建模成本。例如,Kamiya 等^[33]使用 LightGBM 算法和嵌套交叉验证策略优化了 PBPK 模型中吸收速率常数、分布容积和肝内固有清除率这 3 个关键参数的预测,并基于约 14~26 种物理化学性质构建了预测模型。验证结果表明,该模型计算出的大

鼠血浆、肝脏和肾脏中化合物浓度与传统方法相当,显著提高了 PBPK 模型的预测准确性和泛化能力。此外,AI 还被用于 AOP、毒物基因组学、高内涵细胞定量成像等领域的化合物毒性研究^[26]。

相比药物等化合物的毒理学数据,食品中风险物质的毒性数据往往数据量较小且获取成本高,传统的机器学习模型难以构建准确可靠的毒性预测模型,迁移学习的引入可以克服这些挑战。迁移学习是通过迁移不同而相关的元领域的知识,来提高目标学习者在目标领域的学习表现,从而减少构建目标学习模型对大量目标域数据的依赖^[34]。Zhao 等^[35]为了研究蒙古药 Gunqile-7 的毒性,采用一种结合深度卷积神经网络和迁移学习的算法,通过利用预训练的定量毒性预测 QuantitativeTox 模型和数据增强技术,实现了对蒙古药 Gunqile-7 小样本数据的高效毒性预测与分类,二分类任务的准确率达到 0.925,相较于未使用迁移学习的算法至少提高了 9%。

AI 技术还可以用于食品中蛋白质的致敏性评价。传统的致敏性评价方法通常包括体外试验(如 ELISA、免疫印迹等)和体内实验(如动物实验)。这些试验虽准确可靠,但通常需要大量的时间和经济成本。体外试验常常受到试剂和操作复

杂度的限制,而体内实验则涉及伦理和成本问题。AI技术在致敏性评价中的应用主要体现在数据分析和建模方面。Wang等^[36]比较了深度学习自然语言处理模型BERT与LightGBM、XGBoost和随机森林3种集成学习模型对食物蛋白致敏性的预测效果。研究者首先从Allergen Nomenclature、SDAP和NCBI 3个公共数据库收集583个食品致敏原及其蛋白质序列作为正样本,以及600个未报道为致敏原的食物蛋白及其序列作为负样本构建数据集。五折交叉验证结果显示,深度学习模型的AUROC最高可达0.9578,准确率为0.9310,召回率为0.9419,精确率为0.9262,优于传统的集成学习模型。

现如今,昆虫蛋白等替代蛋白的重要性日益凸显,尤其是在满足全球蛋白质需求,推动可持续发展以及应对气候变化方面扮演着关键角色。部分植物蛋白、昆虫蛋白等新型蛋白质来源可能存在不同程度的致敏风险,这不仅涉及已知过敏原的交叉反应,还可能涉及新蛋白质结构引发的未知过敏反应。由于替代蛋白尤其是细胞培养肉等新型蛋白资源的致敏性数据较少,因此AI技术为其致敏性评价提供了新的思路和工具。例如,Westerhout等^[37]采用随机森林模型,开发了一种基于蛋白质理化和生化性质计算变量的模型,用于预测新型蛋白的潜在致敏性。模型对动物、真菌或植物界的任意蛋白质的预测准确率、敏感性和特异性均超过85%,并且成功预测了8种昆虫蛋白质的致敏性,包括之前未在数据库中记录的3种麦蛾幼虫表皮蛋白。

3 快速检测识别元件挖掘

近年来AI已广泛应用于食品检测,包括化学和生物污染物的安全检测、新鲜度等品质检测和真实性鉴别等^[38]。例如,Yang等^[39]提出一种基于机器学习的无损纸显色阵列,该方法利用5层神经网络模型,根据扫描纸显色阵列的颜色变化识别和量化致病菌,可同时检出生菜上大肠杆菌O157:H7和单核细胞增生李斯特菌,准确率高达91%(图3)。Han等^[40]基于深度学习和高光谱成像技术,构建了在像素级别上检测花生中黄曲霉毒素的方法。该研究通过构建一个包含光栅模块、

SCMOS相机和电动位移平台的高光谱成像系统,获取73个花生样本受污染前、后的146个高光谱图像立方体。采用一种像素级光谱重塑图像方法结合卷积神经网络进行机器学习训练,实现了对花生中黄曲霉毒素的检测,整体识别率在95%以上,在核仁级别的识别率也超过90%。Adedeji等^[41]综述基于AI的食物致敏原的无损检测和定量方法,传感器技术和机器学习算法的结合能够快速、准确地检测食物中的致敏原,因无需样品前处理,故检测效率和可及性得到提高。此外,AI辅助的模式识别和数据分析能够显著提升传感器的性能,解决传感器数据中的干扰、校准误差和选择性不足等问题,提高了传感器的分辨率、灵敏度和准确性,最终实现传感器数据的快速分析和处理^[42]。

AI在识别元件的设计和筛选上也发挥了重要作用,然而相关综述较少。本文重点对抗体和适配体两种传统和新型食品安全检测的识别元件进行说明。传统的抗体筛选通常依赖实验室中的高通量筛选技术,通过免疫动物或噬菌体展示技术等方式筛选,从而产生大量的抗体候选分子。传统的实验室筛选技术不仅成本高昂,效率低下,而且涉及动物伦理问题。近年来,单细胞测序,基于X射线晶体和冷冻电子显微镜的基础结构解析,合成生物学和AI技术给抗体设计领域带来了革命性的变革。AI技术使得从头设计抗体成为可能,大大提高了抗体发现的速度和精确度。特别是深度学习模型,能够在庞大的数据集中发现潜在的抗体序列,预测抗体与特定抗原之间的亲和力,从而指导高效的抗体设计。He等^[43]总结了生成扩散模型在抗体设计方面的最新进展,特别是针对从头设计新抗体及优化互补决定区(Complementarity-determining region, CDR)的方法。通过引入属性引导采样、多通道等变分图神经网络和AlphaFold3的不变点注意力等模块,能够生成具有可溶性和更高亲和力、稳定性的抗体序列,而且大量体内实验和体外试验验证了其有效性和实用性。例如,华盛顿大学生物化学教授David Baker团队首次利用RFdiffusion和RoseTTAFold2网络,实现了可与用户指定表位结合的抗体重链可变区的从头设计^[44]。Shanker等^[45]利用结构指导的语言模型ESM-IF1进行蛋白质和抗体复合物的无监

督进化,通过整合蛋白质结构信息来预测并筛选出能显著提高抗体中和能力和亲和力的突变体(图 3)。

适配体是能与靶标分子高亲和力、高特异性结合的单链寡核苷酸,相比抗体具有热稳定性高,便于化学合成与修饰,免疫原性低等优点,在食品安全检测、生物医学等众多领域引起广泛关注。以适配体作为识别元件的检测方法已被用于食源性致病菌、真菌毒素、抗生素、重金属和农残等的检测。适配体常通过指数富集的配基系统进化技术(Systematic evolution of ligands by exponential enrichment, SELEX)筛选得到,然而,由于测序能力和试验条件的限制,传统的 SELEX 方法只能评估有限数量的序列,无法充分探索庞大的核苷酸的理论序列空间,造成很多潜在的高效适配体的遗漏。AI 可以辅助生成和优化高亲和力适配体。

比如, Iwano 等^[46]开发了 RaptGen 模型,用于生成潜在的高效适配体。RaptGen 是一个基于变分自编码器的生成模型,它由一个卷积神经网络编码器和一个基于隐马尔可夫模型的解码器组成。RaptGen 可以从高通量 SELEX 数据中有效地捕捉适配体序列中的局部和全局特征,进而生成新的高亲和力适配体序列。Bashir 等^[47]提出一种机器学习引导的粒子显示方法,用于优化和发现高亲和力适配体。通过粒子显示技术初步筛选适配体,再利用全连接神经网络、卷积神经网络等机器学习模型预测并生成新的适配体序列。生成的适配体具有更高的结合亲和力,其中一些适配体的结合亲和力比原始序列提高了 11 倍,且能够设计出比原始序列短 70% 的截短型适配体,结合亲和力达到 1.5 nmol/L。

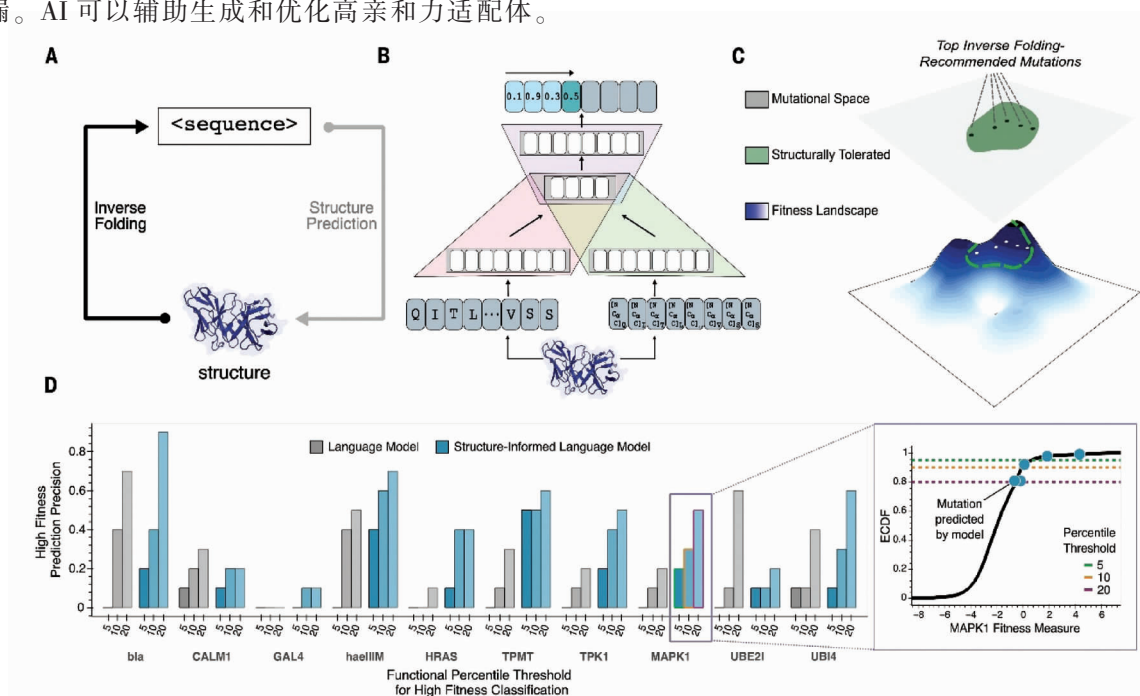


图 3 使用结构引导语言模型来指导多种蛋白质的进化^[27]

Fig.3 Guiding evolution of proteins with a structure-guided language model^[27]

4 生防制剂与高效防控

抗菌肽、降解酶和噬菌体等生防制剂的传统挖掘方法效率低下,成本高昂,结合 AI、生物信息学、分子生物学和合成生物学等工具,可以对这些生防制剂进行挖掘、从头设计、筛选、优化和体外高效表达,探明其与食源性致病菌、真菌毒素等危

害物的特异性识别序列、结合位点、交互方式和裂解途径,从而解析其干预、阻断与控制食源性危害物的关键机制。

4.1 抗菌肽

抗生素耐药性已成为全球公共卫生的重大威胁,抗菌肽通过其广谱抗菌活性,独特的细胞膜破

坏机制,低耐药性风险,协同效应,免疫调节作用以及高度的多样性和可设计性,为应对这一挑战提供了新策略。传统抗菌肽的发现方法依赖于实验室小试,耗时且成本高昂。AI辅助的抗菌肽挖掘方法通过处理大规模微生物组数据,能够快速、准确地预测和筛选出大量潜在的抗菌肽。例如,有研究团队利用机器学习算法从全球微生物组中预测出近百万种新的抗菌肽,利用包含63 410个宏基因组和87 920个原核生物基因组的大型数据集,创建了一个名为AMPSphere的综合目录,该

目录包含了863 498个非冗余肽序列。在测试的100种肽中,有79种在体外显示出活性,其中63种对病原体有效。Torres等^[48]从1 773个人类宏基因组中筛选出具有抗菌特性的候选分子,最终识别出323种由小开放阅读框编码的候选抗菌肽。通过化学合成并测试了78种肽,结果显示70.5%(55种)的肽在体外展示出抗菌活性。这些肽类通过靶向细菌膜杀菌,并且能够相互协同调节肠道共生菌群,显示出重塑微生物群落结构的潜力。

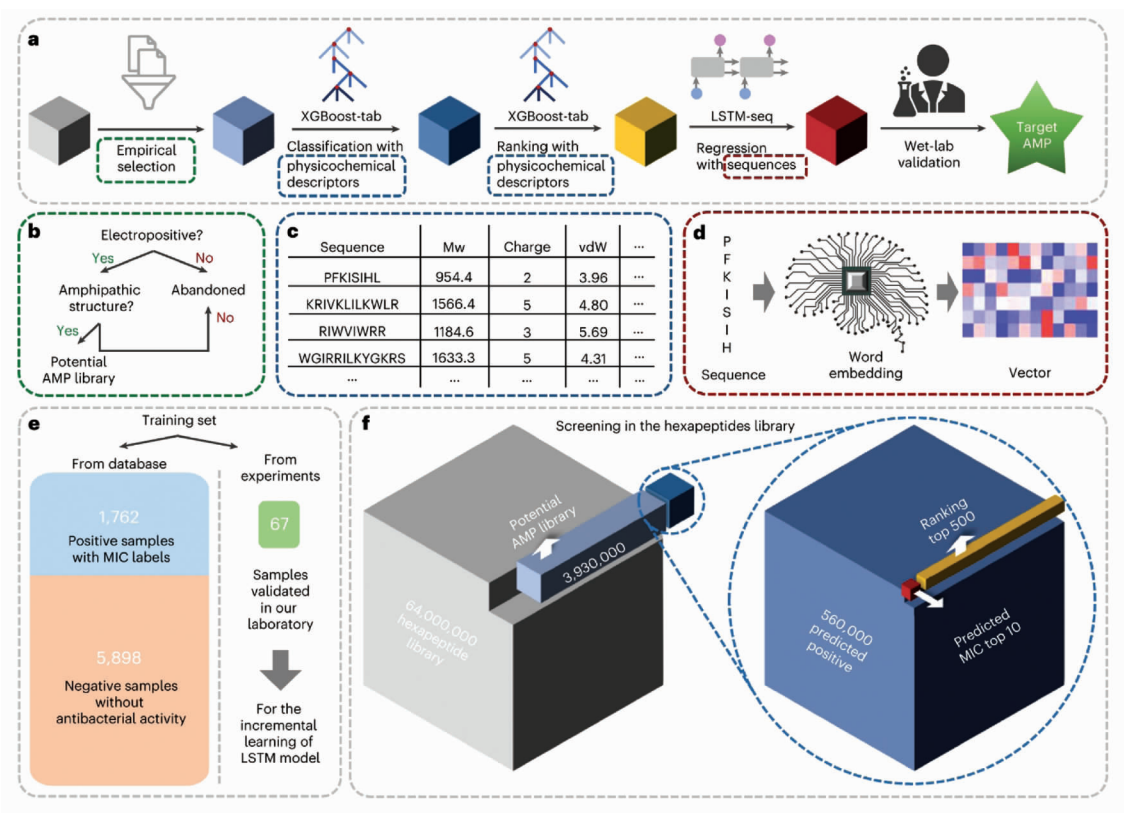


图4 AMP筛选方法概述^[49]

Fig.4 Overview of the antimicrobial peptide screening approach^[49]

大规模微生物组挖掘的方法虽然能够利用自然界中已存在的多样性来识别潜在的抗菌肽,但是存在数据质量和完整性问题,假阳性高以及探索空间受限等挑战,因此采用从头设计方法可以更精准地创造新型抗菌肽,减少假阳性并提高设计效率。Huang等^[49]从APD、DADP、DBAASP、DRAMP和YADAMP等抗菌肽开源数据库筛选了C端酰胺化的肽,并根据其对金黄色葡萄球菌的最小抑菌浓度,构建了包含1 762个阳性和5 898

个阴性抗菌肽的数据集。开发了一个包含分类、排序、回归和增量学习阶段的系统,对由6~9个氨基酸组成的肽的整个虚拟库进行筛选,该库包含数百亿个序列。最终从庞大的候选池中识别出强效的抗菌六肽。经湿试验验证,排名前十的肽中有3种对多种临床分离的多重耐药病原体具有广泛抗菌活性。Li等^[50]利用Uniprot数据库中的大量蛋白质和肽段数据,筛选出长度在10~50之间的约300 000条肽段序列,构建了一个初始训练数据

集。在此基础上,开发了一种基于肽语言的深度生成框架 deepAMP,通过预训练和多次微调策略,结合序列降解方法扩展数据集,以识别具有高效力和广谱抗菌作用的抗菌肽。设计并合成了 29 种候选 AMPs,其中超过 90%的 AMPs 在针对金黄色葡萄球菌等革兰氏阳性菌和铜绿假单胞菌等革兰氏阴性菌的抑制效果上优于渗透肽,抗菌活性最强的抗菌肽效果与 FDA 批准的抗生素相当。

4.2 降解酶

降解酶在生物降解、环境治理和食品加工等领域中扮演着重要的角色。传统的研究方法通常需要耗费大量时间和经费来进行实验室筛选和改造。AI 技术可以辅助降解酶筛选、性能改良以及作用机制研究。通过数据分析和模型建立,AI 可以加速挖掘和优化降解酶的过程,并与合成生物学相结合,推动降解酶在食品生产加工实践中的应用。

AI 可以分析大量的生物信息学数据,从而快速识别具有降解潜力的酶。例如,Lu 等^[51]利用机器学习算法,从大量蛋白质结构中预测并筛选出能够水解聚对苯二甲酸乙二醇酯 [Poly (ethylene terephthalate),PET]的酶,显著提升了筛选效率和准确性(图 5)。此外,通过深度学习模型,如 AlphaFold3 也能够高精度地预测蛋白质结构,并识别可能的活性位点和底物结合模式,为发现降解酶提供了新途径^[52]。另外,在比较基因组学分析方面,AI 还可以识别能够降解特定污染物的微生物,并挖掘对应的降解酶^[53]。

AI 技术在改造和优化酶的降解能力方面也展现出巨大的潜力。借助机器学习模型,研究人员能够预测酶突变对催化性能的影响,并设计出更高效的突变体。例如,Son 等^[54]通过理性蛋白改造和计算设计,成功提升了 PET 水解酶的热稳定性和催化效率。此外,深度学习模型,如 ESM-1v 和 SESNet 等可以预测突变对酶稳定性和活性的影响,从而指导酶的改造。结合结构和计算生物学,AI 技术还能预测酶活性位点的关键氨基酸残基,进一步指导定点突变和定向进化^[55-57]。例如,基于 AI 的蛋白质工程策略,研究人员成功改造了黄曲霉毒素氧化酶,在提高其降解效率与热稳定性方面取得了显著成果^[58]。

AI 还可以模拟酶与底物的相互作用,以揭示酶的催化机制。例如,Yang 等^[59]利用决策树算法结合物理化学特性与结构信息来构建分类器,预测酶的活性。通过交叉验证评估,该模型在预测化合物与糖基转移酶相互作用时准确率达到 90%。Ribeiro 等^[60]开发了 EzMechanism 工具,基于酶机制数据库 Mechanism and Catalytic Site Atlas(M-CSA)的编译规则下自动推断 3D 活性位点和酶反应的催化机制路径。通过对 62 种酶进行验证,EzMechanism 成功找到 28 种完整的酶机制路径。

AI 与合成生物学的结合可以进一步加速降解酶的研发,特别是在设计、优化和应用方面。通过对基因组和代谢途径的工程化改造,合成生物学实现了微生物及其酶系的定向设计与优化。AI 则通过分析大量数据,如基因表达和代谢产物浓度等,加快这些代谢途径的优化过程。除了单个降解酶的优化外,合成生物学还包括构建复杂的酶系和代谢网络。利用 AI 结合合成生物学设计多酶降解系统,并通过调控不同酶的表达水平,成功实现了复杂底物的高效分解,并极大地扩展了降解酶的应用潜力。

4.3 噬菌体

AI 在探索天然噬菌体的过程中也扮演着越来越重要的角色。它不仅帮助研究者从复杂的宏基因组样本中识别噬菌体,还能从其基因组序列中注释出关键的病毒蛋白,准确预测其宿主并挖掘出潜在的抗菌剂等。例如,半监督学习模型 HostG 利用知识图谱和图卷积网络增强了对新型病毒宿主的预测,可以从新分类群中预测宿主^[61]。STEP3 则是利用不同噬菌体基因组的进化特征来实现稳定和准确的预测,解决了基因组注释不佳等问题,提高了噬菌体混合物的质量,实现了更有效的噬菌体治疗^[62]。SCORPION 采用两步特征选择策略构建特征向量和堆叠模型,为仅基于蛋白质一级序列便可准确识别噬菌体病毒粒子蛋白,为其低成本、规模化应用提供了思路^[63]。

新型噬菌体裂解酶也因高效的杀菌能力和高度的宿主专一性,成为新一代抗菌制剂的研究热点。传统的试验性筛选方法工作量大,有研究利用未表征的噬菌体基因组数据,开发了 DeepLysin 软件包,通过机器学习算法从这些基因组中筛选

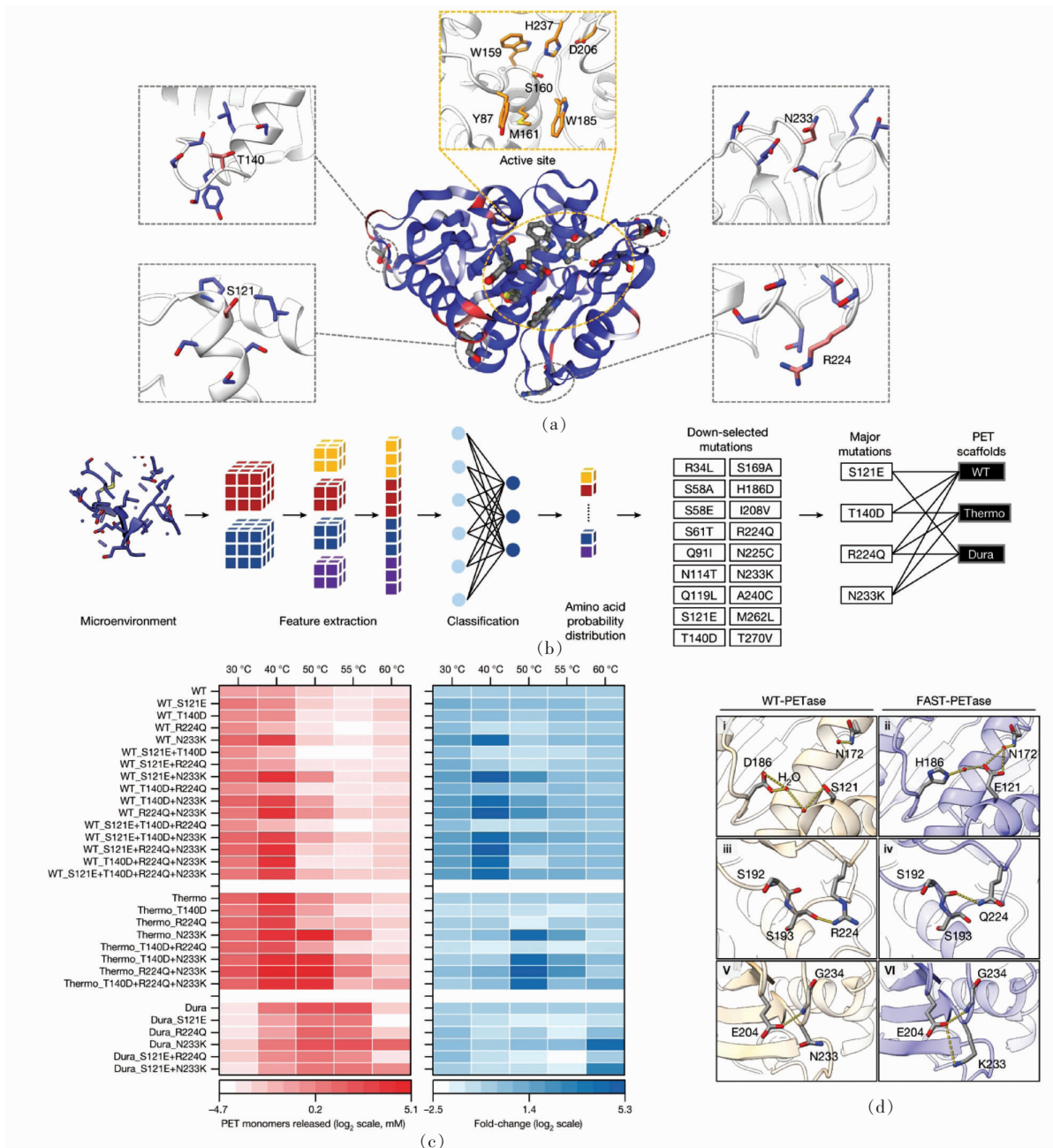


图5 机器学习指导的预测改善了PETase支架的酶性能^[51]

Fig.5 Machine learning guided predictions improve enzyme performance across PETase scaffolds^[51]

潜在的抗菌溶菌酶^[64]。通过对17种随机选择的候选溶菌酶进行试验验证,发现7种具有优异的体外抗菌活性。DeepLysin还可用于原噬菌体预测或挖掘与肽聚糖相互作用的抗菌蛋白。

5 总结与展望

食品安全是全球公共健康的焦点之一。面对食

源性疾病、真菌毒素污染、化学残留等多重挑战,传统手段已难以满足现代社会的需求。AI技术的发展为解决食品安全问题提供了新的思路。AI能够有效地预警食品生产过程中的致病菌、真菌毒素和化学危害物的污染风险,预测新发、突发危害物的毒性,筛选和设计高效识别元件,以及从头设计和优化抗菌肽、降解酶和噬菌体等新型生防制

剂,为食品安全提供了更智能、更有效的保护手段。

面向未来,食品安全问题将更加多样化和复杂化。特别是替代蛋白等新资源食品、跨境食品、电商食品和预制菜等新业态食品,以及极端条件下的太空和深海食品等安全问题,AI技术将持续发挥其巨大潜力。随着大数据以及深度学习、生成式人工智能等技术的发展,AI可以快速识别食品成分并优化食品配方,不仅确保其符合严格的安全标准,还可提升食品的营养价值与口感。同时,建立智能化的监管平台,AI可以实时监控食品生产和流通的各个环节,及时发现并预警潜在的安全隐患,结合计算机视觉技术对食品进行无损检测,确保食品在整个供应链中的安全和品质。AI辅助的新发、突发危害物识别元件和生防制剂筛选、设计和优化,也将进一步提高未来食品安全防控的效率和准确性。未来,AI技术将为应对复杂的食品安全挑战提供更多的创新性解决方案,将在保障全球食品安全方面发挥更加重要的作用。

然而,AI在食品安全领域的应用也面临着诸多挑战。AI技术的应用受到数据共享不足和标准化问题的制约。由于隐私和法律方面的顾虑,数据共享较为困难,因此引发数据孤岛现象。此外,食品供应链数据通常包含多模态信息,如传感器数据、图像数据、文本数据等,且具有时空依赖性。许多农业原材料来自数据意识较为薄弱的发展中或欠发达地区,若无法有效将这些地区纳入AI系统,则算法的性能表现将大打折扣。在极端条件下,数据采集难度大,也将导致AI算法训练数据不足,限制了模型的准确性和泛化性。技术普及、人才短缺、伦理与法律等问题也是不可忽视的重要挑战。未来,政府、企业和科研机构需要共同努力,推动AI技术的健康发展和广泛应用,使其在全球食品安全保障中实现更高效、更全面的应用和创新,助力应对复杂的食品安全挑战。

参 考 文 献

- [1] WHO. Food safety[EB/OL]. (2024-10-04) [2024-10-28]. <https://www.who.int/news-room/fact-sheets/detail/food-safety>.
- [2] UNEP. Food waste index report 2024. Think eat safe: Tracking progress to halve global food waste[EB/OL]. (2024-09-26) [2024-10-28]. <https://go.nature.com/4dD9dHG>.
- [3] ESKOLA M, KOS G, ELLIOTT C T, et al. World-wide contamination of food-crops with mycotoxins: Validity of the widely cited 'FAO estimate' of 25% [J]. *Critical Reviews in Food Science and Nutrition*, 2020, 60(16): 2773-2789.
- [4] 李福君. 我国农户储粮小型粮仓和装具研发应用现状及展望[J]. *粮油食品科技*, 2012, 20(3): 50-52. LI F J. Vista of the farmer's grain storage facilities in China[J]. *Science and Technology of Cereals, Oils and Foods*, 2012, 20(3): 50-52.
- [5] 姚仁朋, 孙玉敬, 赵圆, 等. 机器学习在食品工业中的应用[J]. *中国食品学报*, 2024, 24(1): 349-363. YAO R P, SUN Y J, ZHAO Y, et al. Applications of machine learning in the food industry[J]. *Journal of Chinese Institute of Food Science and Technology*, 2024, 24(1): 349-363.
- [6] MU W, KLETER G A, BOUZEMBRAK Y, et al. Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools[J]. *Comprehensive Reviews in Food Science and Food Safety*, 2024, 23(1): e13296.
- [7] QIAN C, MURPHY S I, ORSI R H, et al. How can AI help improve food safety?[J]. *Annual Review of Food Science and Technology*, 2023, 14: 517-538.
- [8] ESMAEILI R, RAZAVI M A, RAZAVI S H. A step forward in food science, technology and industry using artificial intelligence [J]. *Trends in Food Science & Technology*, 2024, 143: 104286.
- [9] FENG Y, SONI A, BRIGHTWELL G, et al. The potential new microbial hazard monitoring tool in food safety: Integration of metabolomics and artificial intelligence[J]. *Trends in Food Science & Technology*, 2024, 149: 104555.
- [10] SNYDER A B, MARTIN N, WIEDMANN M. Microbial food spoilage: Impact, causative agents and control strategies[J]. *Nature Reviews Microbiology*, 2024, 22(9): 528-542.
- [11] IPCC. Special report: Global warming of 1.5 °C[EB/OL]. (2018-10-06) [2024-10-05]. <https://www.ipcc.org>.

- ch/sr15/?spm=5176.28103460.0.0.297c5d27NN7I4M.
- [12] NEWMAN R, NOY I. The global costs of extreme weather that are attributable to climate change[J]. *Nature Communications*, 2023, 14(1): 6103.
- [13] WHEELER T, VON BRAUN J. Climate change impacts on global food security[J]. *Science*, 2013, 341(6145): 508–513.
- [14] CASU A, CAMARDO LEGGIERI M, TOSCANO P, et al. Changing climate, shifting mycotoxins: A comprehensive review of climate change impact on mycotoxin contamination[J]. *Comprehensive Reviews in Food Science and Food Safety*, 2024, 23(2): e13323.
- [15] CHHAYA R S, O'BRIEN J, CUMMINS E. Feed to fork risk assessment of mycotoxins under climate change influences – recent developments[J]. *Trends in Food Science & Technology*, 2022, 126: 126–141.
- [16] BATTILANI P, TOSCANO P, VAN DER FELS-KLERX H J, et al. Aflatoxin B1 contamination in maize in Europe increases due to climate change[J]. *Scientific Reports*, 2016, 6(1): 24328.
- [17] 蒙继华, 郑贺刚, 王松雪, 等. 基于卫星遥感预测作物真菌毒素的可行性分析[J]. *遥感技术与应用*, 2023, 38(3): 535–543.
- MENG J H, ZHENG H G, WANG S X, et al. The feasibility analysis on satellite data based crop fungal toxin prediction[J]. *Remote Sensing Technology and Application*, 2023, 38(3): 535–543.
- [18] LIU C, MANSTRETTA V, ROSSI V, et al. Comparison of three modelling approaches for predicting deoxynivalenol contamination in winter wheat[J]. *Toxins*, 2018, 10(7): 267.
- [19] WELLER D L, LOVE T M T, BELIAS A, et al. Predictive models may complement or provide an alternative to existing strategies for assessing the enteric pathogen contamination status of northeastern streams used to provide water for produce production[J]. *Frontiers in Sustainable Food Systems*, 2020, 4: 561517.
- [20] ARNING N, SHEPPARD S K, BAYLISS S, et al. Machine learning to predict the source of campylobacteriosis using whole genome data[J]. *PLoS Genetics*, 2021, 17(10): e1009436.
- [21] NDRAHA N, HSIAO H I, HSIEH Y Z, et al. Predictive models for the effect of environmental factors on the abundance of *Vibrio parahaemolyticus* in oyster farms in Taiwan using extreme gradient boosting[J]. *Food Control*, 2021, 130: 108353.
- [22] REVERTER M, SARTER S, CARUSO D, et al. Aquaculture at the crossroads of global warming and antimicrobial resistance[J]. *Nature Communications*, 2020, 11(1): 1870.
- [23] NGUYEN H A, PELEG A Y, SONG J, et al. Predicting *Pseudomonas aeruginosa* drug resistance using artificial intelligence and clinical MALDI-TOF mass spectra[J]. *mSystems*, 2024, 9(9): 18.
- [24] DONG W, HU T, ZHANG Q, et al. Prediction of food safety risk level of wheat in China based on pyraformer neural network model for heavy metal contamination[J]. *Foods*, 2023, 12(9): 1843.
- [25] JIANG B, HE J R, YANG S Q, et al. Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues[J]. *Artificial Intelligence in Agriculture*, 2019, 1: 1–8.
- [26] LIN Z M, CHOU W C. Machine learning and artificial intelligence in toxicological sciences[J]. *Toxicological Sciences*, 2022, 189(1): 7–19.
- [27] SHARMA B, CHENTHAMARAKSHAN V, DHURANDHAR A, et al. Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations[J]. *Scientific Reports*, 2023, 13(1): 4908.
- [28] CIALLELLA H L, ZHU H. Advancing computational toxicology in the big data era by artificial intelligence: Data-driven and mechanism-driven modeling for chemical toxicity[J]. *Chemical Research in Toxicology*, 2019, 32(4): 536–547.
- [29] TROPSHA A, ISAYEV O, VARNEK A, et al. Integrating QSAR modelling and deep learning in drug discovery: The emergence of deep QSAR[J]. *Nature Reviews Drug Discovery*, 2024, 23(2): 141–155.
- [30] PHELPS D W, PARKINSON L V, BOUCHER J M, et al. Per- and polyfluoroalkyl substances in food packaging: Migration, toxicity, and management strategies[J]. *Environmental Science & Technology*, 2024, 58(13): 5670–5684.
- [31] CHENG W X, NG C A. Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list[J]. *Environmental Science & Technology*, 2019, 53(23):

- 13970–13980.
- [32] 周悦, 李潇岚, 程薇, 等. 基于机器学习算法的食品污染物神经毒性预测模型建立[J]. 现代食品科技, 2022, 38(4): 216–223, 306.
ZHOU Y, LI X L, CHENG W, et al. Neurotoxicity prediction model of food contaminants based on machine learning[J]. Modern Food Science and Technology, 2022, 38(4): 216–223, 306.
- [33] KAMIYA Y, HANDA K, MIURA T, et al. *In silico* prediction of input parameters for simplified physiologically based pharmacokinetic models for estimating plasma, liver, and kidney exposures in rats after oral doses of 246 cisparate chemicals[J]. Chemical Research in Toxicology, 2021, 34(2): 507–513.
- [34] ZHUANG F Z, QI Z Y, DUAN K Y, et al. A comprehensive survey on transfer learning[J]. Proceedings of the IEEE, 2021, 109(1): 43–76.
- [35] ZHAO H K, QIU S, BAI M R, et al. Toxicity prediction and classification of Gunqile-7 with small sample based on transfer learning method[J]. Computers in Biology and Medicine, 2024, 173: 108348.
- [36] WANG L Y, NIU D T, ZHAO X J, et al. A comparative analysis of novel deep learning and ensemble learning models to predict the allergenicity of food proteins[J]. Foods, 2021, 10(4): 809.
- [37] WESTERHOUT J, KRONE T, SNIPPE A, et al. Allergenicity prediction of novel and modified proteins: Not a mission impossible! Development of a Random Forest allergenicity prediction model[J]. Regulatory Toxicology and Pharmacology, 2019, 107: 104422.
- [38] 丁浩晗, 谢祯奇, 沈嵩, 等. 人工智能在食品检测中的应用[J/OL]. 食品科学技术学报: 1–13 (2024–10–08)[2024–10–28]. <http://kns.cnki.net/kcms/detail/10.1151.ts.20240930.1724.002.html>.
DENG H H, XIE Z Q, SHEN S, et al. Application of artificial intelligence in food testing[J/OL]. Journal of Food Science and Technology: 1–13 (2024–10–08)[2024–10–28]. <http://kns.cnki.net/kcms/detail/10.1151.ts.20240930.1724.002.html>.
- [39] YANG M Y, LIU X B, LUO Y G, et al. Machine learning-enabled non-destructive paper chromogenic array detection of multiplexed viable pathogens on food[J]. Nature Food, 2021, 2(2): 110–117.
- [40] HAN Z Z, GAO J Y. Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging[J]. Computers and Electronics in Agriculture, 2019, 164: 104888.
- [41] ADEDEJI A A, PRIYESH P V, ODUGBEMI A A. The magnitude and impact of food allergens and the potential of AI-based non-destructive testing methods in their detection and quantification[J]. Foods, 2024, 13(7): 994.
- [42] ZHOU Z Z, XU T L, ZHANG X J. Empowerment of AI algorithms in biochemical sensors[J]. TrAC Trends in Analytical Chemistry, 2024, 173: 117613.
- [43] HE X H, LI J R, XU J, et al. AI-driven antibody design with generative diffusion models: Current insights and future directions[J/OL]. Acta Pharmacologica Sinica: 1–10 (2024–09–30)[2024–10–28]. <https://doi.org/10.1038/s41401-024-01380-y>.
- [44] BENNETT N R, WATSON J L, RAGOTTE R J, et al. Atomically accurate de novo design of single-domain antibodies[J/OL]. bioRxiv: 1–30 (2024–03–18)[2024–10–28]. <https://doi.org/10.1101/2024.03.14.585103>.
- [45] SHANKER V R, BRUUN T U J, HIE B L, et al. Unsupervised evolution of protein and antibody complexes with a structure-informed language model[J]. Science, 2024, 385(6704): 46–53.
- [46] IWANO N, ADACHI T, AOKI K, et al. Generative aptamer discovery using RaptGen[J]. Nature Computational Science, 2022, 2(6): 378–386.
- [47] BASHIR A, YANG Q, WANG J P, et al. Machine learning guided aptamer refinement and discovery[J]. Nature Communications, 2021, 12(1): 2366.
- [48] TORRES M D T, BROOKS E F, CESARO A, et al. Mining human microbiomes reveals an untapped source of peptide antibiotics[J]. Cell, 2024, 187(19): 5453–67.e15.
- [49] HUANG J J, XU Y C, XUE Y F, et al. Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences[J]. Nature Biomedical Engineering, 2023, 7(6): 797–810.
- [50] LI T T, REN X B, LUO X L, et al. A foundation model identifies broad-spectrum antimicrobial peptides against drug-resistant bacterial infection[J]. Nature Communications, 2024, 15(1): 7538.
- [51] LU H Y, DIAZ D J, CZARNECKI N J, et al. Machine learning-aided engineering of hydrolases for PET depolymerization[J]. Nature, 2022, 604(7907): 662–667.

- [52] ABRAMSON J, ADLER J, DUNGER J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. *Nature*, 2024, 630(8016): 493–500.
- [53] 向霞, 朱恩恒, 韩楠玉. 三种主要真菌毒素及其毒素降解酶的研究进展[J]. *生物技术通报*, 2024, 40(1): 45–56.
- XIANG X, ZHU E H, HAN N Y. Research progress in three major mycotoxins and their toxin-degrading enzymes[J]. *Biotechnology Bulletin*, 2024, 40(1): 45–56.
- [54] SON H F, CHO I, JOO S, et al. Rational protein engineering of thermo-stable PETase from *Ideonella sakaiensis* for highly efficient PET degradation [J]. *ACS Catalysis*, 2019, 9(4): 3519–3529.
- [55] CUI Y L, CHEN Y C, SUN J Y, et al. Computational redesign of a hydrolase for nearly complete PET depolymerization at industrially relevant high-solids loading[J]. *Nature Communications*, 2024, 15(1): 1417.
- [56] YEH A H-W, NORN C, KIPNIS Y, et al. De novo design of luciferases using deep learning[J]. *Nature*, 2023, 614(7949): 774–780.
- [57] RIVES A, MEIER J, SERCU T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences[J]. *Proceedings of the National Academy of Sciences*, 2021, 118(15): e2016239118.
- [58] HU Y M, LI H, MIN J, et al. Crystal structure and biochemical analysis of the specialized deoxyribovalenol-detoxifying glyoxalase SPG from *Gossypium hirsutum*[J]. *International Journal of Biological Macromolecules*, 2022, 200: 388–396.
- [59] YANG M, FEHL C, LEES K V, et al. Functional and informatics analysis enables glycosyltransferase activity prediction [J]. *Nature Chemical Biology*, 2018, 14(12): 1109–1117.
- [60] RIBEIRO A J M, RIZIOTIS I G, TYZACK J D, et al. EzMechanism: An automated tool to propose catalytic mechanisms of enzyme reactions[J]. *Nature Methods*, 2023, 20(10): 1516–1522.
- [61] SHANG J Y, SUN Y N. Predicting the hosts of prokaryotic viruses using GCN-based semi-supervised learning[J]. *BMC Biology*, 2021, 19(1): 250.
- [62] THUNG T Y, WHITE M E, DAI W, et al. Component parts of bacteriophage virions accurately defined by a machine-learning approach built on evolutionary features[J]. *mSystems*, 2021, 6(3): 17.
- [63] AHMAD S, CHAROENKWAN P, QUINN J M W, et al. SCORPION is a stacking-based ensemble learning framework for accurate prediction of phage virion proteins[J]. *Scientific Reports*, 2022, 12(1): 4106.
- [64] ZHANG Y, LI R Z, ZOU G, et al. Discovery of antimicrobial lysins from the “dark matter” of uncharacterized phages using artificial intelligence [J]. *Advanced Science*, 2024, 11(32): 2470193.

Advances in Artificial Intelligence-Assisted Proactive Prevention and Control of Food Safety

Sheng Lina¹, Ji Jian¹, Song Xiaoning², Wu Xiaojun², Sun Xiulan^{1*}

¹*School of Food Science and Technology, Jiangnan University, Wuxi 214122, Jiangsu*

²*School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, Jiangsu*

Abstract Food safety is crucial to global public health. In recent years, new hazards have arisen from changing consumption patterns and novel food sources, alongside traditional threats like foodborne pathogens, mycotoxins, and chemical residues. The rapid advancement of artificial intelligence (AI) technology provides innovative solutions to address these challenges. This article summarizes the applications of AI technology in the proactive prevention and control of food safety, including risk warning, toxicity prediction, rapid detection, and efficient prevention and control strategies. AI integrates meteorological statistics, mechanistic models, and machine learning algorithms to achieve early warning of risk factors such as foodborne pathogens, mycotoxins, pesticides, and heavy metals in food. AI also assists traditional toxicological models and incorporates transfer learning to facilitate toxicity prediction and risk assessment of new hazards. Besides rapid detection for food safety and quality, AI also plays a role in high-throughput design and screening of both

traditional and novel recognition elements such as antibodies and aptamers. In the realm of proactive prevention and control, AI combines with bioinformatics, molecular biology, and synthetic biology to effectively predict and screen antimicrobial peptides, degrading enzymes, and bacteriophages, revealing their antimicrobial and degradation mechanisms. However, the application of AI in food safety still faces challenges such as insufficient data sharing, standardization, and handling of multimodal data. With the advancements of technology and evolvement of data sharing mechanisms, AI is expected to play an increasingly important role in ensuring global food safety and addressing complex and evolving food safety issues.

Keywords artificial intelligence; food safety; risk early warning; toxicity prediction; rapid detection; prevention and control