

机器学习方法在小麦表型和基因型研究领域的应用进展

张志贤¹, 孙萌浩¹, 陈 博¹, 李志豪¹, 苏培森², 孟宪勇¹, 黄思罗³, 柳平增¹, 颜 君^{1*}

(1 山东农业大学信息科学与工程学院, 农业农村部黄淮海智慧农业技术重点实验室, 泰安 271018; 2 聊城大学农学与农业工程学院, 聊城 252000; 3 华中科技大学生命科学与技术学院, 武汉 430074)

摘 要: 随着高通量测序技术、遥感技术、机器学习技术的迅速发展, 小麦 (*Triticum aestivum* L.) 研究正从传统的表型评估向数据驱动的精准分析逐渐转变。机器学习作为处理海量数据、发现潜在模式的核心技术, 已在小麦研究的多个领域得到广泛应用。本文系统总结了机器学习在小麦四大核心领域的研究进展: 生长状态与生理特性监测、生物胁迫与非生物胁迫识别、产量与品质预测、抗病抗逆基因挖掘。通过多源数据融合与算法创新, 机器学习算法实现了对小麦生理参数的无损监测、胁迫的早期诊断、产量的精准预测以及育种候选基因的挖掘。然而, 当前研究仍面临数据异质性、模型泛化能力不足、可解释性缺失等挑战。未来基于机器学习的小麦相关研究需重点突破多模态数据协同学习、机理-数据双驱动建模、轻量化部署三大方向, 构建“从基因到田间”的智能化育种与管理体系, 为全球粮食安全与农业可持续发展提供核心驱动力。

关键词: 机器学习; 小麦生长监测; 胁迫识别; 产量预测; 抗病抗逆基因挖掘

中图分类号: Q37; S512.1; TP181 **文献标识码:** A

Advances in the application of machine learning methods in wheat phenotypic and genotypic research

ZHANG Zhi-Xian¹, SUN Meng-Hao¹, CHEN Bo¹, LI Zhi-Hao¹, SU Pei-Sen²,
MENG Xian-Yong¹, HUANG Si-Luo³, LIU Ping-Zeng¹, YAN Jun^{1*}

(1 Key Laboratory of Huang-Huai-Hai Smart Agricultural Technology of the Ministry of Agriculture and Rural Affairs, College of Information Science and Engineering, Shandong Agricultural University, Taian 271018, China; 2 College of Agronomy and Agricultural Engineering, Liaocheng University, Liaocheng 252000, China; 3 College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: With the rapid advancement of high-throughput sequencing, remote sensing, and machine learning technologies, wheat (*Triticum aestivum* L.) research is gradually transitioning from traditional phenotypic evaluation to data-driven precision analysis. As a core technology for processing massive high-dimensional data and uncovering underlying complex patterns, machine learning has effectively addressed key research challenges such as wheat's large and intricate genome (≈ 16 Gb) and the nonlinear relationships between genotype and phenotype. It has been deeply applied in multiple fields of wheat research and demonstrated revolutionary value. This study systematically summarizes the research progress of machine learning in four core areas of wheat research over the past decade (2015–2025): monitoring of growth status and physiological traits, identification of biotic and abiotic stresses, prediction of yield and quality, and mining of disease-resistance and stress-tolerance genes. Through multi-source data fusion (including UAV remote sensing, hyperspectral imaging, 3D point clouds, multi-omics data, etc.) and algorithmic innovation (from traditional machine learning to deep learning model optimization), machine learning has achieved non-destructive and high-precision monitoring of wheat physiological parameters, early and

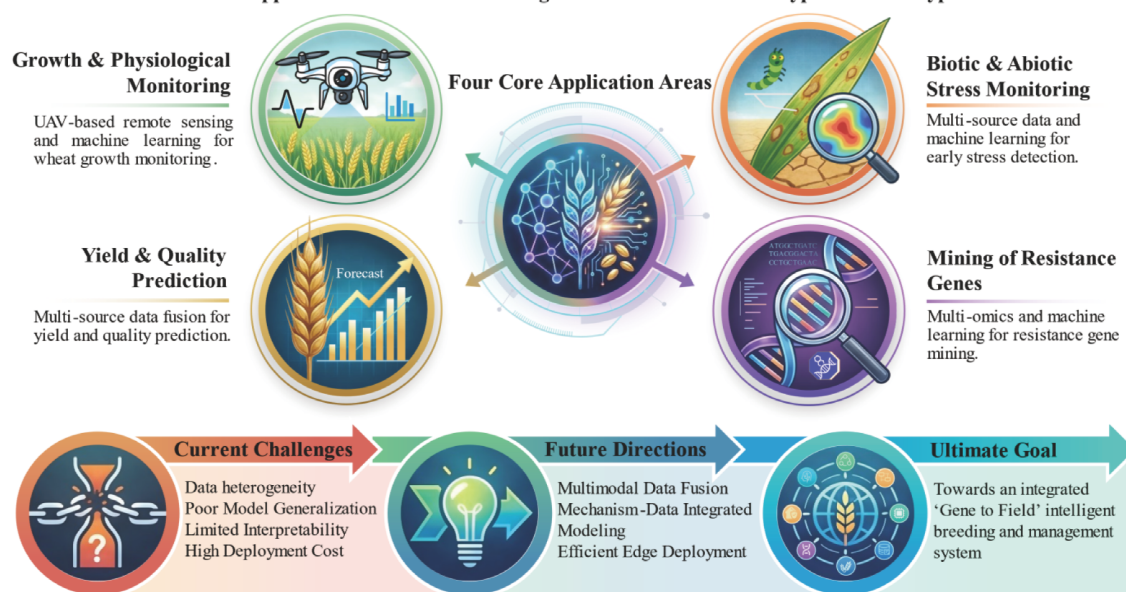
收稿日期: 2026-01-06; 修回日期: 2026-02-21

基金项目: 山东省高等学校“青创人才引进计划”项目(基于机器学习的农业生物信息学创新团队)

*通信作者: E-mail: xinsinian2006@163.com

accurate diagnosis of stresses, quantitative prediction of yield and quality, as well as efficient screening of candidate breeding genes. However, current research still faces prominent challenges: strong data heterogeneity and lack of unified standards, insufficient cross-environment generalization ability and poor interpretability of models, as well as high computational power requirements and high costs for field deployment. Future wheat-related research based on machine learning should focus on breaking through three key directions: advancing the collaborative learning of "space-air-ground" integrated multi-modal data, constructing a mechanism-data dual-driven modeling framework, and developing lightweight models and edge computing deployment solutions. Ultimately, by integrating technological achievements, an integrated intelligent breeding and management system "from gene to field" will be built, providing a core driving force for global food security and sustainable agricultural development.

Advances in the Application of Machine Learning Methods in Wheat Phenotypic and Genotypic Research



Key words: machine learning; wheat growth monitoring; stress identification; yield prediction; mining of disease-resistance and stress-tolerance genes

小麦作为全球最重要的粮食作物之一,其稳产高产对保障粮食安全具有重要的战略意义。根据《中国2023统计年鉴》数据,2023年中国小麦种植面积达 2.363×10^7 公顷,约占全国粮食作物总种植面积的19.86%,凸显了其在我国农业生产中的核心地位。在高通量测序、遥感技术、人工智能(artificial intelligence, AI)快速发展的背景下,小麦研究正由传统表型评估逐渐向数据驱动的精准分析转变。然而,小麦基因组规模庞大(约16 Gb),结构复杂,存在数据体量大、特征维度高、非线性关系复杂等挑战^[1]。

机器学习(machine learning, ML)作为人工智能的核心分支,能够从大规模数据中提取特征并挖掘潜在规律,为小麦研究提供了革命性的技术工具。近年来,机器学习方法已在小麦研究的多个领域得到广泛应用,涵盖小麦生长状况监测、胁迫识别与预测^[2]、产量和质量研究^[3]、基因挖掘^[4]等领

域,显著提升了研究效率和预测准确性。

本文系统梳理了过去十年(2015—2025年)小麦研究中机器学习的应用进展,重点围绕小麦生长状况及生理特性监测、胁迫识别与预测、产量和品质研究、抗病抗逆基因挖掘这四个核心领域展开综述。通过对相关研究的分析,本文总结当前研究热点,分析相关研究趋势,指出存在问题,为小麦机器学习领域的研究人员提供一定的参考^[5]。

1 小麦生长状态与生理特性监测

精准无损地进行生理参数监测是智慧农业决策的核心基础。当前小麦生理参数监测研究主要聚焦于叶绿素含量、株高、水分状况、生物量及氮磷钾含量等关键指标的采集分析,通过机器学习算法构建高精度的反演模型,为精准管理和智能育种提供决策支持。图1总结了机器学习在小麦生长状态与生理特性监测研究中的关键步骤,一般分为4步:(1)数

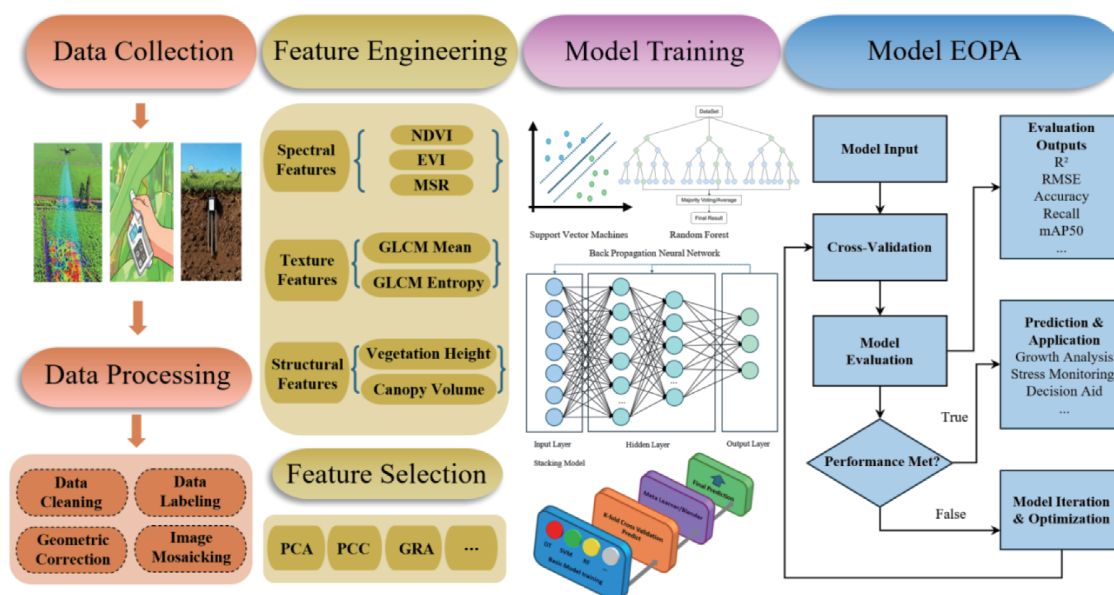


图1 机器学习在小麦生长状态和生理特性监测中的关键步骤

Figure 1 Key steps of machine learning in monitoring wheat growth status and physiological characteristics
 Model EOPA指模型评估(Evaluation)、优化(Optimization)、预测(Prediction)和应用(Application)

据收集(如无人机(unmanned aerial vehicle, UAV)高光谱遥感数据、叶绿素检测数据、土壤传感监测数据等)和数据预处理;(2)特征工程和特征筛选:从原始数据中提取出与小麦有关的光谱特征、纹理特征、结构特征等,并从这些特征中进一步筛选出最具相关性和区分度的核心特征;(3)模型训练:基于筛选得到的高相关性小麦生长特征数据集,选取适配的机器学习算法开展模型训练;(4)评估应用:选取准确率、召回率、mAP50等评价指标对模型性能进行量化评估,并基于评估结果对模型进行迭代改进,然后将训练好的模型用于实际场景,为小麦管理提供精准的决策支持。

1.1 叶绿素含量监测

在小麦叶绿素含量监测研究中,机器学习技术已成为实现快速无损预测的核心工具,通过多源数据融合显著提升了估算精度,呈现出低成本、高效率的趋势。2022年,Wang等^[6]使用UAV获取正常灌溉和干旱胁迫下小麦在抽穗期、开花期、灌浆期等多个时期的多光谱图像,使用随机森林(random forest, RF)算法优选14个植被指数;接着采用9种机器学习模型预测叶绿素含量,结果显示带交叉验证的岭回归(ridge regression with cross-validation, RidgeCV)模型在正常灌溉抽穗期表现最佳($r=0.63$),支持向量机(support vector machine, SVM)模型在干旱胁迫下精

度最高($r=0.63$)。2024年,Singh等^[7]利用LemnaTec-Scanalyzer3D平台采集小麦RGB图像,计算39种RGB植被指数,并通过HTPmod开源平台的16种机器学习算法预测叶绿素含量,结果显示双向循环神经网络(bidirectional recurrent neural network, BRNN)模型达到 $R^2=0.71$ 的高精度,为低成本监测提供了可行方案。2024年,王志业^[8]通过粒子群算法(particle swarm optimization, PSO)优化支持向量回归模型(support vector regression, SVR),使PSO-SVR模型测试集 R^2 达0.791,显著提升了小麦叶绿素含量的估测精度。当前小麦叶绿素监测研究呈现出从单一传感器到多源数据融合、从简单回归到复杂机器学习乃至深度学习(deep learning, DL)模型的明显趋势。其中,低成本的RGB(red green blue)方案潜力巨大,但需克服环境泛化难题。

1.2 株高量化技术

株高是衡量小麦生长状况和生物量的关键表型参数,传统人工测量方法效率低下且易引入误差。机器学习技术已成为实现高效无损监测小麦株高的核心工具,推动其从传统人工测量向自动化高精度分析转变。2021年,付晶波等^[9]使用基于掩码的区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)模型对盆栽小麦进行实例分割,结合图像处理提取株高,将模型输出的株

高数据与人工测量数据对比, R^2 达0.983,这表明了深度学习在植物株高提取上的高精度优势。2024年,Bello等^[10]综述了3D点云深度学习技术,分析了点云数据的不规则性、非结构化和无序性等挑战,指出通过采样、分组和映射函数(如PointNet架构)可直接处理原始点云数据,这为株高的三维数字化分析提供了理论支撑。在此基础上,Dagar等^[11]开发了基于RGB-D相机(Intel RealSense D455)的实用化流程,通过多视角图像采集生成高精度点云,然后使用人工神经网络(artificial neural network, ANN)实现株高自动化测算;该方法在80株小麦样本上验证显示,ANN模型的性能(RMSE=13.70)显著优于传统方法(如线性回归)。这些研究虽取得了显著进展,但仍面临点云数据处理复杂、模型泛化能力不足等挑战,未来需开发自动化算法并整合多模态数据以提升田间适用性。

1.3 小麦水分含量监测

在小麦水分含量监测研究中,机器学习技术已成为实现精准无损预测的核心工具,通过多源数据融合与算法创新显著提升了水分胁迫诊断的准确性;水分状况监测关键指标包括冠层水分含量(canopy water content, CWC)、植株水分含量(plant water content, PWC)、冠层等效水厚度(canopy equivalent water thickness, CEWT)、叶片含水量(leaf water content, LWC)等。2024年,Zununjan等^[12]引入分数阶导数(fractional order derivative, FOD)预处理高光谱数据,并结合K近邻(K-nearest neighbors, KNN)、SVM等算法构建LWC预测模型,最佳模型(FWBI-3BI-0.8阶) $R^2=0.86$, RMSE=2.11%。同年,Yadav等^[13]基于冠层温度、气象、土壤参数,采用SVR等4种模型预测作物水分胁迫指数(crop water stress index, CWSI),得出SVR性能最好,在特定输入组合下实现 $R^2=0.997$,这项研究展示了ML在预测CWSI方面的巨大应用前景。2023年,Zhuang等^[14]融合了连续小波变换(coupling continuous wavelet, CWT)与ML改进了冬小麦水分状况预测;通过RF和SVM预测CWC和PWC,验证集 R^2 分别达0.92和0.93,利用特征重要性分析(如WF2318, 4对CWC贡献度36%)增强模型的可解释性。这些研究推动了小麦水分监测从单一指标向多参数融合逐步演进。

1.4 生物量估算

在小麦生长监测研究中,地上生物量是指小麦

植株地上部分的有机物质重量,包括鲜重和干重,是评估作物生长状况、光合效率、最终产量的关键农艺参数。近年来,机器学习技术结合UAV遥感技术已成为高通量估算小麦生物量的有效工具。2022年,Schreiber等^[15]基于RGB图像和ANN、CNN模型,对小麦生物量进行回归预测,分别实现了 $R^2=0.906$ 和0.907的高精度验证,表明该方法可用于精准农业中地上生物量的时空变化预测。同年,Sharma等^[16]利用UAV多光谱图像衍生的VIs(vegetation indices, 植被指数)数据,使用偏最小二乘法(partial least squares, PLS)、SVM、RF、ANN算法,得到在美国南达科他州Beresford地点的抽穗后阶段生物量估算,PLS、RF、SVM模型表现最佳, R^2 为0.70~0.71,而ANN模型稍差($R^2\approx 0.63$)。2023年,Zhu等^[17]整合了不同飞行高度的UAV获取的光谱、纹理、冠层表面等模型数据,使用RF、SVR、深度卷积神经网络(deep convolutional neural network, DCNN)模型,实现了小麦生物量的精准估算,发现多尺度数据整合显著提升了模型精度(DCNN模型的 R^2 最高达0.84)。2024年,Kc等^[18]进一步引入方差膨胀因子(variance inflation factor, VIF)特征选择方法,通过剔除冗余特征以降低计算成本并提升预测精度^[19],从而应用于小麦生物量的预测;首先对多光谱图像处理得到了13个植被指数(VIs),然后从中筛选出6个最优VIs集,采用极端梯度提升(eXtreme gradient boosting, XGBoost)、RF等6种模型,使用筛选后6个最优VIs训练得到的模型精度整体优于或近似于基于全部13个指数训练的模式。

近年来,多源数据融合与机理模型结合进一步提升了生物量估算精度。2022年,Li等^[20]建立了全生长季节冬小麦生物量估算模型CBA-Wheat(crop biomass algorithm for wheat);该模型将物候信息(扎多克生育期量表等,一种精确编码作物生长阶段的体系)作为校正因子,动态调整植被指数与生物量之间的线性模型参数,解决了生物量估算模型在时空可转移性和全生育期普适性方面的关键限制;但该模型在卫星遥感应用中依赖大量地面实测数据调参,限制了其推广使用。2023年,王士俊等^[21]针对CBA-Wheat的上述问题,采用遗传算法(genetic algorithm, GA)对模型参数进行优化,耦合多光谱遥感数据优化作物生长模型;优化后的模型实现了对冬小麦全生育期生物量的高精度估算, R^2 达0.92。

2025年,Chen等^[22]提出相对日的概念来取代传统的ZS(Zadoks scale,一种用于描述作物生长阶段的物候数据),从而优化CBA-Wheat模型;改进后的模型对各个物候阶段的响应敏感,且 R^2 达0.82。

1.5 氮磷钾含量监测

在小麦氮、磷、钾含量监测领域,机器学习方法与无人机多光谱遥感(unmanned aerial vehicle multispectral remote sensing technology, UAV-MS RS)、无人机高光谱遥感(unmanned aerial vehicle hyperspectral remote sensing technology, UAV-HS RS)的结合已成为研究热点。通过融合光谱特征与纹理信息,构建高效的小麦微量元素含量预测模型,

为精准农业提供重要的技术支撑。表1总结了部分相关研究成果。

2 小麦生物胁迫与非生物胁迫监测

2.1 生物胁迫监测与识别

近年来,为应对小麦生产中的主要生物胁迫,研究者们综合利用多源遥感影像、高光谱数据、田间采集图像等多模态数据,并结合ML和DL等多种算法,在病虫害的识别、严重度评估、发生预测等方面取得了重要进展。这些研究为监测和早期预警小麦病虫害提供了有力的技术支撑。我们总结了近期相关的研究进展,见表2。

表1 机器学习监测小麦的氮、磷、钾含量

Table 1 Machine learning monitors wheat nitrogen, phosphorus and potassium content

研究主题	研究元素	数据类型	机器学习方法	典型性能	参考文献
小麦氮状态监测	N	多光谱图像	基于5种U-Net	Attention R2U-Net, 准确率99.96%	[23]
估计小麦不同叶位氮含量	N	高光谱反射率	PLSR, SVM, RF, MLR	MLR, R^2 达0.89	[24]
小麦氮含量评估	N	高光谱图像	ANN	ANN, R^2 达0.86	[25]
冬小麦叶片氮含量估计	N	多光谱图像	SVR, RF, RR, KNN, Voting, Stacking	Stacking, R^2 达0.82	[26]
小麦叶片氮含量估计	N	UAV-MS、气象、土壤数据	ANN, ALR, MLR	ANN, R^2 达0.745	[27]
预测磷对春小麦产量影响	P	气候指标、区域特征、磷施用数据	ANN	ANN, R^2 达0.87	[28]
诊断冬小麦氮营养指数	N	UAV-HS、地面光谱数据	LR	LR, R^2 达0.75	[29]
监测冬小麦叶片氮含量	N	UAV-HS、纹理特征	RR, PLSR, SV, RF	RR, R^2 达0.87	[30]
小麦钾含量预测	K	光谱数据	PLS, RF	PLS, R^2 达0.74	[31]

注:“典型性能”所列指标源于各原始文献在其特定实验设置下报告的最优结果,用于展示不同技术路径所能达到的典型性能水平,因研究目标与条件差异,不宜进行直接横向比较;PLSR(partial least squares regression, 偏最小二乘回归);MLR(multiple linear regression, 多元线性回归);RR(ridge regression, 岭回归);Voting(投票);Stacking(堆叠)

表2 机器学习在小麦生物胁迫监测与识别中的应用

Table 2 Application of machine learning in wheat biological stress monitoring and identification

监测目标	胁迫类型	机器学习方法	数据来源	参考文献
小麦蚜虫发生范围预测	小麦蚜虫	RVM, SVM, LR	光学数据/热红外数据/实地调查数据	[32]
小麦螨虫自动识别和计数	小麦螨虫	CNN, ZFNet, VGG16	田间数字图像	[33]
小麦条锈病严重度评估	WSR	K-means, SC, SVM, RF, KNN	病害图像	[34]
小麦害虫自动识别和分类	多种虫害	4种ResNet	IP102数据集害虫图像	[35]
小麦叶部病害图像识别	WSR/WLR/PM	SVM, KNN, GNB, MLP, SSL	多设备病害图像	[36]
小麦叶枯病疾病症状识别	STB	RF	近端传感器数据	[37]
预测小麦赤霉病发生	FHB	logistic-KNN	遥感数据/气象数据	[38]
监测小麦穗部赤霉病严重度	FHB	K-NN, RF, SVM	高光谱反射/叶绿素成像/高通量表型	[39]
监测小麦白粉病严重度	PM	PLSR, SVR, RFR	高光谱反射数据	[40]

注:WSR(wheat stripe rust, 小麦条锈病);WLR(wheat leaf rust, 小麦叶锈病);STB(*Septoria tritici* blotch, 小麦叶枯病);FHB(*Fusarium head blight*, 小麦赤霉病);PM(powdery mildew, 小麦白粉病);RVM(relevance vector machine, 相关向量机);LR(logistic regression, 逻辑回归);ZFNet(Zeiler& Fergus network, 泽勒-弗格斯网络);VGG16(visual geometry group 16, 视觉几何组16层网络);K-means(K-means clustering algorithm, K均值聚类);SC(spectral clustering, 谱聚类);ResNet(residual network, 残差网络);GBN(gaussian naive bayes, 高斯朴素贝叶斯);MLP(multi-layer perceptron, 多层感知机);SSL(self-supervised learning, 自监督学习);PLSR(partial least squares regression, 偏最小二乘回归);RFR(random forest regression, 随机森林回归)

2.2 非生物胁迫监测与诊断

非生物胁迫监测与诊断技术针对干旱、高温、涝渍、盐胁迫、重金属胁迫等胁迫类型,已发展出基于高光谱、多光谱、机器学习的系统性方法,这些技术在原理创新、模型优化(通过特征工程、超参数调优与集成学习等方法,旨在提升模型的诊断准确率、泛化能力及计算效率)、田间验证中展现出独特优势。

其中,高温胁迫和水分胁迫(包括干旱、涝渍)虽然表现出不同的环境压力形式,但二者存在一定的关联性,常常协同影响小麦的生长周期。高温胁迫通过促进蒸腾作用加速植株水分流失,从而加剧干旱胁迫;而水分胁迫中的涝渍情况则通过土壤缺氧抑制根系呼吸和水分吸收,进一步干扰植株的热调节能力,形成“水分-热”复合胁迫效应^[41]。表3总结了部分相关研究成果。未来研究需进一步探索多胁迫交互下的动态响应模型,结合机理驱动框架提升预测的泛化能力。

土壤盐渍化是影响小麦生长的关键非生物胁迫因素。融合遥感技术和机器学习方法为实现大范围、无损监测种植小麦的盐胁迫提供了新方案。

2020年,Qi等^[50]构建了“卫星-UAV-地面”协同反演土壤盐分的方法;该研究对比反向传播神经网络(back propagation neural network,BPNN)、SVM、RF模型后,发现基于UAV影像的RF模型对土壤盐分反演精度最高;将该结果与Sentinel-2A卫星影像结合进行协同反演,显著提高了监测准确性,解决了卫星成像质量低和UAV观测范围有限的问题。2022年,Fu等^[51]开发了小麦种子活力评估系统(wheat seed vigour assessment system,WSVAS),通过比较三种目标检测模型(VGG16+Faster R-CNN、ResNet50+Faster R-CNN、YOLOv4),确定使用YOLOv4监测不同盐胁迫下小麦种子的发芽曲线为最佳模型(MAP=97.59%);该研究结果表明,WSVAS可以量化盐胁迫引起的植物胁迫反应,为小麦耐盐育种提供了量化性指标。同年,Kecoglu等^[52]采用拉曼光谱结合多种机器学习回归模型(包括高斯过程回归、SVM等),对小麦叶片的盐胁迫程度进行精准数值化测定;其中,有理二次高斯过程回归模型表现最佳(R^2 达0.923~0.933);该模型可以识别在盐胁迫初始阶段所引起的小麦代谢物含量变化

表3 机器学习在小麦温度胁迫与水分胁迫研究中的应用

Table 3 Research on machine learning in wheat under temperature stress and water stress

胁迫类型	监测技术	关键诊断参数	机器学习方法	技术路径/优势	文献参考
高温+干旱	高通量表型分析	Tr、BA、cytINV、vacINV、AC	SVM、LR、BM	多源传感整合,机理阐释	[42]
干旱	高光谱遥感、CWT	CWC、PWC、CEWT	MLR、RF、SVM	小波-机器学习耦合	[14]
干旱	叶绿素荧光成像	PSII效率、纹理特征、颜色特征	RF、KNN	叶绿素荧光图像处理,高准确率(91.164%)	[43]
干旱+热胁迫	生物电信号测量	ESF、振幅、熵、信号强度	RF、KNN、SVM	生物电传感,早期胁迫识别	[44]
干旱+热胁迫	遥感技术	Tc、NDVI、ChlF、WC	CNN、RF	多尺度遥感-多组学整合,动态机制解析与预测	[45]
热胁迫	无人机多光谱成像	SPAD、SDR、EVI、NDVI	DT、RF、SVM、LR	UAV多光谱,SDR驱动的高精度耐热筛选	[46]
水分胁迫	无人机多光谱遥感	SMC、FMC、PMC、AGB	集成模型	多特征融合与WE-stacking集成,高精度($R^2=0.884$)	[47]
水分胁迫	红外温度计等	Tc、Ta、VPD、Rn、U、SD	SVR、ANN、RFR、MLR	SVR高精度领先($R^2=0.997$)	[13]
水分胁迫	热红外/RGB成像	Tc、RWC、SMC、RH	10种机器学习方法	深度学习多模态图像分析,优势:高精度胁迫分类(准确率98.4%)	[48]
水分胁迫	无人机多光谱遥感	gs、SWSI、NDVI、纹理特征	SEL、RF、SVM、XGB、MLR	多特征融合与堆叠集成学习,预测精度显著提升(R^2 提高6.25%~14.63%)	[49]

注:Tr(transpiration,蒸腾作用)、BA(biomass accumulation,生物量积累)、cytINV(细胞壁转化酶)、vacINV(液泡转化酶)、AC(antioxidant capacity,抗氧化能力)、PSII效率(photosystem II efficiency,光系统II的光化学效率)、ESF(electrical signal features,电信号特征)、冠层温度(Tc)、NDVI(normalized difference vegetation index,归一化植被指数)、WC(water contents,水分含量)、SPAD(叶绿素相对含量指数)、SDR(SPAD decline rate,SPAD下降率)、EVI(enhanced vegetation index,增强型植被指数)、SMC(soil moisture content,土壤水分含量)、FMC(fuel moisture content,燃料水分含量)、Ta(空气温度)、VPD(vapor pressure deficit,蒸汽压亏缺)、Rn(净太阳辐射)、U(风速)、SD(soil moisture depletion,土壤水分耗竭)、RWC(relative water content,相对水分含量)、RH(relative humidity,相对湿度)、gs(气孔导度)、SWSI(soil water stress index,土壤水分胁迫指数)、SEL(stacking ensemble learning,堆叠集成学习)

(如纤维素、果胶、氨基酸含量等)。这些研究表明,多平台数据融合与高维光谱技术,结合机器学习模型,在精准、早期量化作物胁迫程度方面潜力巨大,为小麦抗逆栽培与育种筛选提供了高效无损的技术支持。

重金属胁迫(如镉、砷、铜、锌污染)是影响小麦生长与粮食安全的重要非生物胁迫因素。采用高光谱遥感技术结合ML能够非破坏性地获取小麦的光谱信息,通过分析光谱特征变化间接估测土壤的重金属含量。例如,2017年,Wang等^[53]通过田间试验发现,小麦在铜、锌胁迫下冠层光谱数据在可见光波段的反射率随重金属浓度增加而升高,而在近红外波段结果则相反;他们利用红边归一化植被指数(NDVI705)成功识别了小麦重金属胁迫临界浓度。2023年,Zhai等^[54]综述了可见/近红外光谱和高光谱成像在植物重金属胁迫表型分析中的应用,强调机器学习在提高检测精度和实现高通量筛选方面具有优势。这些研究为利用机器学习实现小麦重金属胁迫的快速无损监测提供了理论基础。牛硕等^[55]构建了RF、BPNN、Freundlich-type转移方程模型,预测了小麦镉富集因子(BCF-Cd,即小麦籽粒镉含量与土壤镉含量的比值,用于量化镉在土壤-小麦系统的迁移能力),结果表明RF模型预测精度最优,土壤磷含量和锌含量是影响小麦镉富集因子的关键因素。同年,Zhong等^[56]对小麦叶片高光谱数据进行平滑降噪、噪声波段剔除及7种数学变换预处理,然后用遗传算法优化的PLSR建立镉和砷含量的估测模型,其决定系数分别达0.87和0.91,显著优于传统方法。这些案例表明,机器学习不仅提升了重金属胁迫预测的准确性,还通过特征重要性分析揭示了关键环境因子,为小麦的安全生产提供了决策支持。

3 小麦产量与品质研究

3.1 产量估算与预测

在小麦产量预测研究中,准确预估收获量是优化农业管理和保障粮食安全的核心环节。早期研究侧重于通过经典机器学习模型的比较与优化来提升预测精度。2021年,Bali和Singla^[57]针对印度旁遮普邦小麦产区,利用长短期记忆网络(long short-term memory, LSTM)来解决循环神经网络(recurrent neural network, RNN)模型固有的梯度消失问题,构建了RNN-LSTM小麦产量预测模型。

基于1970—2012年(43年)的气象与产量数据,该模型性能显著优于ANN、RF和MLR,其均方根误差最小(147.12 kg/ha),为小麦产量的精准预测提供了有效的技术方案。2022年,魏萌秀^[58]以山西运城市种植小麦为研究对象预测产量,结合气象因子与物候期数据,对比了BPNN、极限学习机(extreme learning machine, ELM)、径向基函数(radial basis function, RBF)模型,结果表明RBF模型预测精度最优;然后利用PSO算法对三种模型进行优化,其中PSO-RBF模型优化效果最为显著,预测误差降至1.354%。江斌超^[59]针对黄淮海地区的小麦,采用HP滤波法将产量分离为趋势产量和气候产量;然后使用8种机器学习模型对小麦产量趋势进行预测(模型含:决策树回归(decision tree regression, DTR)、线性回归(linear regression, LR)、K近邻回归(K-nearest neighbor regression, KNR)、RFR、AdaBoost回归(adaboost regression, ABR)、梯度提升回归(gradient boosting regression, GBR)、Bagging回归(bagging regression, BR)、极端随机树回归(extreme random tree regression, ETR)),发现河南、山东、河北三省的最优预测模型不相同,但所有模型 R^2 均大于0.9,预测效果优异;接着使用自回归积分移动平均模型(autoregressive integrated moving average, ARIMA)预测小麦气候产量,最后通过叠加趋势产量和气候产量的预测结果,实现了对黄淮海区域小麦总产量的精准预测。2023年,王奎^[60]使用BP、Elman、SVR模型预测河南省小麦产量,采用PSO和GA算法优化三种模型;结果显示PSO-Elman模型MAPE仅为0.35%,该研究验证了智能优化算法在减小局部极小值风险方面的优势。这些研究验证了智能优化算法与产量分解思想在提升模型精度方面的有效性。

随着技术的发展,小麦产量预测研究重点逐渐转向多元数据融合与集成学习,以进一步提升预测鲁棒性。2022年,Misra等^[61]提出Yield-SpikeSegNet方法,通过收缩-扩张网络检测小麦麦穗,结合ANN和SVR估计产量;该方法在高通量和非破坏性的小麦表型分析领域取得了重大进展。2023年,Fei等^[62]通过融合UAV多传感器数据并结合一种称为Stacking的集成学习方法,实现了对小麦产量的高精度预测;相比于单一传感器数据,多传感器数据融合能显著提升预测准确性;另外,集成学习模型性能

($R^2=0.692$) 优于单一ML模型。2024年,齐浩等^[63]收集了小麦灌浆期的UAV高光谱和产量数据,然后使用9种机器学习模型(RR、PLS、MLR、RF、GBR、XGB、高斯过程回归(Gaussian process regression, GPR)、SVR、KNN)进行产量估测,结果表明GBR在单一模型中表现最好;随后,他们采用Stacking集成学习方法,将KNN、RR、SVR、GBR多个单一模型的预测结果作为输入,训练出一个更高层次的模型进行最终决策,结果发现 R^2 提高了1.39%;该研究表明多模型集成能有效克服单一算法的局限性。同年,Yang等^[64]使用UAV多传感器数据(RGB、多光谱、纹理、热红外),结合三种集成方法(简单平均集成方法、叠加法、特征加权法)进行建模预测,结果表明简单平均集成方法(对叠加法 and 特征加权法的结果取算术平均值)性能高于叠加法 and 特征加权法;该研究表明多源数据融合提高了预测鲁棒性。这些研究表明,多源数据融合与集成学习是提升产量预测鲁棒性的关键方向,但跨区域泛化能力和数据标准化仍是未来需要突破的难点。

综上分析可见,在小麦产量预测研究中,传统机器学习模型(如RBF)通常结构简单、训练高效,在中小规模数据集上表现稳定,如魏萌秀^[58]的研究中RBF模型经PSO优化后误差可降至1.354%,显示了其在快速建模中的潜力。然而,此类模型在捕捉复杂非线性关系方面存在局限性;相比之下,深度学习模型(如LSTM)擅长处理时序依赖的高维数据。例如,Bali和Singla^[57]利用LSTM处理了积累43年的气象数据,取得了较低的RMSE(147.12 kg/ha),凸显了其在长期复杂序列中的预测优势,但其对数据量和计算资源要求较高,且模型可解释性较差。近年来,集成学习方法(如Stacking)成为提升模型鲁棒性的主流方向。Fei等^[62]和齐浩等^[63]的研究均表明,通过整合多个基学习器的优势,集成模型的预测精度(R^2)可比单一最优模型进一步提升,有效克服了单一算法的局限性,但代价是模型复杂度和计算成本增加。因此,方法的选择需权衡任务复杂度、数据规模和可解释性的需求。

3.2 品质与加工特性研究

小麦品质特性研究主要涵盖营养成分检测(如蛋白质、湿面筋、微量元素)、多指标综合评价等内容。近年来,高光谱成像、近红外光谱、多传感器融合等技术取得突破,结合机器学习算法,显著提升

了小麦品质检测的精度、效率与可视化能力。2021年,张磊^[65]系统地将机器学习应用于小麦品质的多指标预测(脂肪酸、发芽率、电导率等),提出长短期记忆-生成对抗网络(LSTM generative adversarial network, LSTM-GAN)时序模型预测小麦品质,准确度较传统LSTM提升了9.745%;为实现多指标联合分析,他们进一步构建了Broad-AdaBoost模型,通过集成学习提升了预测稳定性,同时构建的半监督分类模型(DP-TWSVM)实现了品质状态的智能识别,准确率达到95.83%,为多指标综合评价奠定基础。2024年,Zhang等^[66]从高光谱和RGB传感器数据提取原始反射率(original reflectance, OR)、小波特征(wavelet feature, WF)、显色指数(color index, CI)等特征,将Pearson-CARS-VIF特征选择方法与4种机器学习算法(BPNN、PLSR、ELM、RFR)结合,对小麦籽粒的蛋白质含量和湿面筋含量进行估测;结果表明基于多模态数据融合的RF模型实现了最高的预测精度,且小波特征的贡献度最大。同年,Shi等^[67]采用VIS-NIR高光谱成像技术与3种ML模型(PLSR、最小绝对收缩与选择算子回归(least absolute shrinkage and selection operator, LASSO)、逐步线性回归(stepwise linear regression, SLR)),对小麦籽粒的营养素进行无损定量分析,发现基于一阶导数处理后的高光谱数据(dA)的SLR模型(SLR-dA)可以可靠预测220种营养素;同时该研究使用改进的pix2pix生成对抗网络,实现了小麦籽粒营养素空间分布的可视化。2025年,宋茂兴等^[68]开发了基于近红外光谱与堆叠集成学习的小麦种子蛋白质含量的无损检测方法,该研究对比了5种机器学习模型(LR、PR、RFR、PLSR、SVR)和堆叠模型(分别采用SWG和LSG进行分组)后发现,基于LSG的堆叠模型能够有效整合光谱的全局信息,通过集成多个子模型减轻共线性干扰,为小麦种子蛋白质含量的快速、无损、高精度检测提供了一种优越的算法解决方案。

4 小麦抗病抗逆基因的挖掘

在作物育种中,发掘抗病抗逆基因是开发抗性品种的核心环节。近年来,机器学习算法与生物信息学的融合显著提高了基因识别的精度和效率。通过整合多组学数据(如转录组、基因组)和先进算法,研究者能够从海量数据中筛选出小麦抗病抗逆关键

基因^[69-71],并验证其功能。2018年,Gonzalez-Camacho等^[72]在基因组选择(genomic selection, GS)框架下针对小麦锈病抗性基因预测开展研究,系统比较了RRBLUP、Bayes LASSO等参数线性模型与机器学习方法(如RFR、SVR);结果显示RFR在多数数据集上表现优异,能有效捕获大效应QTL(quantitative trait locus,数量性状基因座)和epistasis(上位性,不同QTL间的非加性相互作用)之间的相互作用,凸显了ML在基因挖掘中的潜力。2021年,Sandhu等^[73]针对小麦品质性状(如面粉产量、籽粒蛋白含量),评估了深度学习模型(CNN、MLP)与传统方法在多环境数据中的表现;研究发现深度学习在籽粒蛋白含量等低遗传力性状上优势显著,能通过自动特征提取处理非线性关系和“基因-环境”互动,进而为早期世代筛选提供理论支持。2022年,Ibrahim等^[74]为挖掘小麦抗生物胁迫相关分子标记,整合了差异表达基因(differentially expressed genes, DEGs)分析和ML技术,基于6种小麦生物胁迫相关基因表达数据集(GEO数据库),鉴定出24 152个差异表达基因;同时采用可解释机器学习模型(extra-tree regressor、LIME)进行生物胁迫关键响应基因筛选。2024年,Zhu等^[75]针对谷子(*Setaria italica*)干旱响应基因挖掘展开研究,先通过机器学习分析转录组数据得到46个候选基因,再用主成分分析法从中筛选出12个重要干旱响应基因,并通过qPCR验证其响应干旱胁迫功能。2025年,He等^[4]结合机器学习与生物信息学分析识别小麦干旱应激响应基因,他们收集了172个小麦RNA-seq样本,通过序列比对和定量表达分析鉴定出16 754个DEGs;再通过加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)识别出与干旱应激相关的模块(如METurquoise模块);接着在对比多种机器学习模型后,发现RF表现最佳(准确率0.889);随后应用改进的RF-Boruta算法筛选出51个干旱基因,模型准确率提升至0.942, AUC达0.978;最后通过基因富集分析揭示这些基因参与苯丙烷代谢等过程,证实了其在干旱响应中的功能。图2总结了抗病抗逆基因的提取流程:(1)数据收集与预处理:从NCBI/CNCB等获取RNA-seq数据,经质量评估、过滤及比对计算基因表达量;(2)差异表达与功能分析:鉴定DEGs并开展GO/KEGG富集;(3)WGCNA分析:基于DEGs构建基因共表达网络并筛

选核心基因;(4)GWAS分析:结合基因型与表型数据开展关联分析以挖掘候选基因;(5)机器学习建模:实现数据处理、特征选择及模型构建与评估;(6)候选基因验证:通过多维度数据集及实验手段验证基因可靠性^[4]

5 结论与展望

本文系统总结了机器学习技术在小麦研究中的四大核心领域(生长监测、胁迫识别、产量品质预测、抗病抗逆基因挖掘)的研究进展。研究表明,机器学习已从辅助工具发展为驱动小麦研究范式变革的核心力量,通过多源数据融合(如高光谱遥感、三维点云、多组学数据)与算法创新(从ML到DL),实现了对小麦生理参数的无损监测、胁迫早期诊断、产量精准预测、基因型-表型关联解析。

然而,当前相关研究仍面临多重瓶颈:(1)数据层面,不同来源(遥感、组学、田间实测等)数据的异质性导致融合效率低,缺乏统一标准与共享平台;具体到小麦研究,其庞大的基因组(≈ 16 Gb)和复杂的基因-表型非线性关系^[76],使得多模态数据(如高光谱影像与转录组数据)的维度与尺度差异巨大。(2)模型层面,现有算法多依赖特定环境下的标注数据,泛化能力不足,难以适应小麦不同产区(如黄淮海与长江中下游麦区)的环境变异^[77];同时,模型的“黑箱”特性导致可解释性不足,预测结果难以关联农学机理^[78];(3)落地层面,深度学习算力要求较高,田间部署成本高昂,限制了其在小麦大规模生产中的应用^[79]。

未来基于机器学习的小麦相关研究需聚焦三大核心方向突破:

(1)推进多模态数据的协同学习^[62,80-82],其核心是发展异质混合专家模型(如STF-MoE)^[80]与集成学习框架(如Stacking回归)^[62]的融合范式,该范式旨在通过动态门控机制,为不同类型的数据匹配适宜的专业化子模块,再通过元学习器整合结果,并借助Transformer(一种深度学习模型)的自注意力机制与多实例学习,动态融合“天-空-地”一体化传感数据。具体而言,可借鉴STF-MoE模型的异质专家池设计,采用结构差异化的神经网络分别处理遥感光谱数据、环境变量及田间表型数据^[80],同时利用集成学习中的二级元学习器(用于整合多个基础模型预测结果的专门模型,核心是通过学习基础模型的优劣表现,动

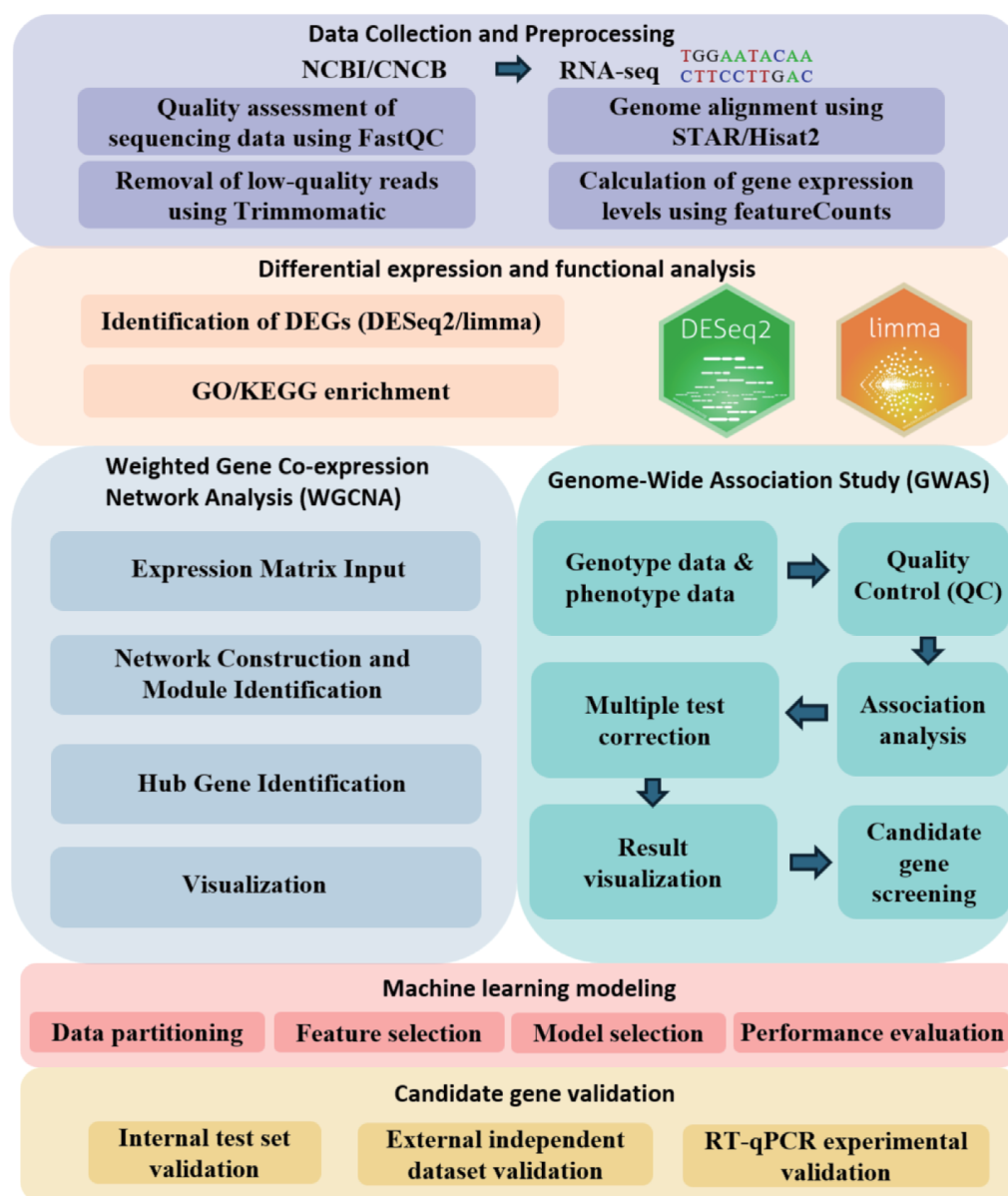


图 2 小麦抗病抗逆基因提取流程

本图展示了从数据预处理、差异表达分析、共表达网络构建、全基因组关联分析到机器学习建模及候选基因验证的完整技术路线。

Figure 2 Extraction process of wheat disease and stress resistance genes

This figure illustrates the complete technical pipeline from data preprocessing, differential expression analysis, co-expression network construction, genome-wide association study to machine learning modeling and candidate gene validation.

态优化多源数据对应的权重分配)优化多源数据权重分配^[62]。针对小麦基因组庞大和“基因-表型”非线性关系复杂的特点,可引入多模态大语言模型(multimodal large language model,MLLM)技术,通过检索增强生成(retrieval-augmented generation,RAG)和人类反馈强化学习(reinforcement learning from human feedback,RLHF)机制对齐跨尺度数据^[82]。例如,在融合高光谱影像与转录组数据时,采用

Transformer-LSTM混合架构捕获长期依赖关系,同时利用专家网络专业化处理不同模态数据的分布特征,最终通过自适应门控网络实现特征贡献度的动态加权,从而显著提升多模态数据的融合效率^[80,83]。

(2)构建“机理-数据双驱动”的建模方法^[84]。在泛化能力方面,现有算法高度依赖特定环境下的标注数据,导致模型在跨生态区(如黄淮海麦区与

长江中下游麦区)应用时性能显著下降。究其原因,小麦生长受气候、土壤、管理措施等多因素交互影响,单一的环境数据难以覆盖全局变异^[85]。为此,可引入迁移学习与领域自适应技术^[77],通过预训练模型学习源区域(如黄淮海麦区)的基础特征,再针对目标区域(如长江中下游麦区)数据进行少量微调,以降低环境差异的干扰。例如,Chen等^[22]提出的基于物候信息校正的地上生物量预测模型CBA-Wheat,通过整合多地点(北京小汤山国家精准农业研究中心、北京郊区(39.60°N-40.30°N,116.47°N-116.97°E)、山东德州禹城)高光谱植被指数与标准化物候参数(relative day of year,RDOY)增强时空适应性,模型 $R^2=0.82$,跨年份、跨生态区验证 $R^2 \geq 0.82$,优于PLSR($R^2=0.78$)和RF($R^2=0.73$),提升了时空适应性,有效降低了环境差异带来的泛化误差。另外,元学习框架能利用历史任务经验快速适配新环境,提升模型在数据稀缺区域的稳健性。例如,模型无关的元学习(model-agnostic meta-learning,MAML)框架的核心机制是在多样化历史任务(如不同生态区、不同年份的小麦数据)上进行元训练,优化初始参数以获得“快速学习”能力;应用于数据稀缺新环境时,仅需少量本地样本微调,即可快速适配,相比传统预训练微调,该方法能避免少样本过拟合,提升泛化性与稳健性^[86]。

在可解释性方面,“黑箱”模型难以将预测结果与农学机理(如作物生理过程、“基因-环境”互作)直接关联,这限制了模型在育种筛选和栽培管理决策中的可信度和应用价值。例如,在小麦产量预测中,如果模型无法解释为何某个基因型在特定环境下表现优异,育种者就无法据此优化亲本选配^[87]。为解决这一问题,需大力发展可解释人工智能(explainable artificial intelligence,XAI)方法,如SHAP(shapley additive explanations)和LIME(local interpretable model-agnostic explanations)等工具。这些方法能解析模型决策逻辑:SHAP通过博弈论概念量化每个特征对预测结果的贡献度,并提供全局特征重要性排名和局部个体样本解释;LIME则通过构建局部替代模型,近似刻画黑箱模型在特定预测点的决策行为,揭示影响单次决策的关键因素^[88]。例如,在Yu等^[87]开发的EXGEP框架中,SHAP被用于识别影响玉米产量的关键基因型和环境特征(如2米高处特定湿度和土壤有机质),并可

可视化其交互作用。Ryo等^[89]则通过LIME分析免耕(不翻动土壤的耕作方式)对玉米产量的影响,发现局部决策中蒸发散(指土壤蒸发和植物蒸腾的总和,即农田中水分以水汽形式散失到大气中的两种过程的统称)等环境因素的主导作用。通过应用可解释人工智能技术,模型不仅能提升预测精度,还能输出农学意义明确的解释(如特定基因型在高温下的适应性),从而辅助育种者制定数据驱动的跨环境育种策略。

(3)落地轻量化部署^[90-92]技术,开发面向田间的轻量化模型^[93]与边缘计算方案^[94]来有效降低硬件门槛与应用成本。具体而言,轻量化部署技术聚焦于模型压缩与结构优化,例如Qiu等^[90]的YOLO-SDL模型引入ShuffleNetV2架构与深度可分离卷积(depthwise separable convolution,DWConv),参数量降至1.76 M,计算量仅5.1 GFLOPs(giga floating-point operations per second,十亿次浮点运算每秒,是衡量模型计算“工作量”的核心指标),在复杂田间环境中保持高精度(mAP50达96.5%)。Qu等^[91]基于YOLOv11n改进的Star-YOLO模型,采用StarNet模块与移动倒置瓶颈卷积(mobile inverted bottleneck convolution,MBConv)轻量化设计,参数量仅2.2 M,在嵌入式设备上实现27 FPS(frames per second,帧/秒)的实时检测。边缘计算方案则通过分布式架构降低云端依赖,例如Dey等^[94]的Fly框架采用5G femtolet(一种兼具通信与高存储、计算能力的小型基站)作为边缘节点,本地化处理传感器数据,结合Bi-LSTM算法进行产量预测,延迟降低约25%,仅将结果上传至云,减少了网络流量与存储开销。这些技术通过模型压缩与边缘协同,使小麦检测、害虫识别等任务能在低成本嵌入式系统(如树莓派、Jetson Nano)上运行,硬件门槛从专业GPU平台降低至普通物联网设备,为深度学习在小麦生产中的大规模普及提供了可行路径。

在此基础上,相关研究可借鉴植物基因组研究跨物种知识迁移与预训练大模型(如WheatGP)的思路,整合三大方向技术成果构建全链条智能框架,实现从“单点优化”到“全链条智能”的跃迁^[95]。最终跨过“突破数据壁垒”“融合农学机理”“降低部署门槛”三个门槛,构建“从基因到田间”的小麦智能化育种与管理体系,为全球粮食安全与农业可持续发展提供核心驱动力。

参考文献

- [1] Tang T, Tian S, Wang H, et al. Wheat-RegNet: An encyclopedia of common wheat hierarchical regulatory networks. *Mol Plant*, 2023, 16: 318–21.
- [2] Ullah N, Ahmad B, Khan A, et al. Attention-guided wheat disease recognition network through multi-scale feature optimization. *ICCK Transact Sensing Commun Control*, 2025, 2: 11–24.
- [3] 李远斌. 融合多模态数据和机器学习算法的山东冬小麦产量预测研究[D]. 哈尔滨: 哈尔滨师范大学, 2023.
- [4] Li Y. Yield prediction of winter wheat in Shandong province: integrating multimodal data and machine learning algorithms[D]. Harbin: Harbin Normal University, 2023.
- [5] He J, Cui B, Liu P, et al. Utilizing machine learning and bioinformatics analysis to identify drought stress responsive genes in wheat (*Triticum aestivum* L.). *Front Sustain Food Syst*, 2025, 9: 1612009.
- [6] Rasheed A, Liu J, Appels R, et al. Mobilizing *Triticeae* diversity from gene banks to farmer's field. *Mol Plant*, 2025, 18: 566–9.
- [7] Wang W, Cheng Y, Ren Y, et al. Prediction of chlorophyll content in multi-temporal winter wheat based on multispectral and machine learning. *Front Plant Sci*, 2022, 13: 896408.
- [8] Singh B, Elangovan A, Kumar S, et al. Non-destructive chlorophyll prediction by machine learning techniques using RGB-based vegetation indices in wheat. *Plant Physiol Rep*, 2024, 29: 836–47.
- [9] 王志业. 基于无人机多光谱遥感和机器学习的小麦生长参数估算研究[D]. 郑州: 河南财经政法大学, 2024.
- [10] Wang Z. Estimation of wheat growth parameters based on UAV multispectral remote sensing and machine learning[D]. Zhengzhou: Henan University of Economics and Law, 2024.
- [11] 付晶波, 施家伟, 张俊, 等. 基于Mask R-CNN的盆栽小麦单片叶长和株高提取研究. *吉林农业大学学报*, 2021, 43: 163–70.
- [12] Fu J, Shi J, Zhang J, et al. Extraction of single leaf length and plant height of potted wheat based on Mask R-CNN. *J Jilin Agric Univ*, 2021, 43: 163–70.
- [13] Bello SA, Yu S, Wang C, et al. Deep learning on 3D point clouds. *Remote Sens*, 2020, 12: 1729.
- [14] Dagar P, Arora A, Ray M, et al. High-resolution reconstruction of images for estimation of plant height in wheat using RGB-D camera and machine learning approaches. *Curr Sci*, 2024, 127: 1440–6.
- [15] Zununjan Z, Turghan MA, Sattar M, et al. Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices. *Plant Methods*, 2024, 20: 97.
- [16] Yadav A, Narakala LM, Upreti H, et al. Evaluation of crop water stress index of wheat by using machine learning models. *Environ Monit Assess*, 2024, 196: 970.
- [17] Zhuang T, Zhang Y, Li D, et al. Coupling continuous wavelet transform with machine learning to improve water status prediction in winter wheat. *Precis Agric*, 2023, 24: 2171–99.
- [18] Schreiber LV, Atkinson Amorim JG, Guimarães L, et al. Above-ground biomass wheat estimation: Deep learning with UAV-based RGB images. *Appl Artif Intell*, 2022, 36: e2055392.
- [19] Sharma P, Leigh L, Chang J, et al. Above-ground biomass estimation in oats using UAV remote sensing and machine learning. *Sensors (Basel)*, 2022, 22: 2055392.
- [20] Zhu W, Rezaei EE, Nouri H, et al. UAV flight height impacts on wheat biomass estimation via machine and deep learning. *IEEE J Sel Top Appl Earth Observ Remote Sens*, 2023, 16: 7471–85.
- [21] Kc K, Romanko M, Perrault A, et al. On-farm cereal rye biomass estimation using machine learning on images from an unmanned aerial system. *Precis Agric*, 2024, 25: 2198–225.
- [22] Jiao R, Nguyen BH, Xue B, et al. A survey on evolutionary multiobjective feature selection in classification: Approaches, applications, and challenges. *IEEE Transact Evol Comput*, 2024, 28: 1156–76.
- [23] Li Z, Zhao Y, Taylor J, et al. Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sens Environ*, 2022, 273: 112967.
- [24] 王士俊, 刘苗, 赵钰, 等. 多源遥感数据耦合CBA-Wheat模型的冬小麦生物量估算研究. *遥感学报*, 2024, 28: 3123–35.
- [25] Wang S, Liu M, Zhao Y, et al. Estimating winter wheat biomass by coupling the CBA-Wheat model and multispectral remote sensing. *Nat Remote Sens Bull*, 2024, 28: 3123–35.
- [26] Chen J, Yu J, Meng Y, et al. Adjusted CBA-wheat model for predicting aboveground biomass in winter wheat from hyperspectral data. *Field Crops Res*, 2025, 333: 110060.
- [27] Blekanov I, Molin A, Zhang D, et al. Monitoring of grain crops nitrogen status from UAV multispectral images coupled with deep learning approaches. *Comput Electron Agric*, 2023, 212: 108047.

- [24] Ma C, Zhai L, Li C, et al. Hyperspectral estimation of nitrogen content in different leaf positions of wheat using machine learning models. *Appl Sci*, 2022, 12: 7427.
- [25] Sahoo RN, Rejith RG, Gakhar S, et al. Drone remote sensing of wheat N using hyperspectral sensor and machine learning. *Precis Agric*, 2023, 25: 704–28.
- [26] Han Y, Zhang J, Bai Y, et al. Ensemble learning-driven and UAV multispectral analysis for estimating the leaf nitrogen content in winter wheat. *Agronomy*, 2025, 15: 1621.
- [27] Fu Z, Yu S, Zhang J, et al. Combining UAV multispectral imagery and ecological factors to estimate leaf nitrogen and grain protein content of wheat. *Eur J Agron*, 2022, 132: 126405.
- [28] Sharipova S, Akanova A, Ospanova N. Development of a neural network model for training data on the effects of phosphorus on spring wheat growth. *East Eur J Enterp Technol*, 2023, 6: 32–8.
- [29] Huangfu L, Jiao J, Chen Z, et al. Diagnosis of winter wheat nitrogen status using unmanned aerial vehicle-based hyperspectral remote sensing. *Appl Sci*, 2025, 15: 1869.
- [30] Zhang J, Cheng T, Shi L, et al. Combining spectral and texture features of UAV hyperspectral images for leaf nitrogen content monitoring in winter wheat. *Int J Remote Sens*, 2022, 43: 2335–56.
- [31] Yang T, Lu J, Liao F, et al. Retrieving potassium levels in wheat blades using normalised spectra. *Int J Appl Earth Observ Geoinform*, 2021, 102: 102412.
- [32] 唐翠翠, 黄文江, 罗菊花, 等. 基于相关向量机的冬小麦蚜虫遥感预测. *农业工程学报*, 2015, 31: 201–7.
Tang C, Huang W, Luo J, et al. Remote sensing prediction of aphids in winter wheat based on relevance vector machine. *Transact Chin Soc Agric Engin*, 2015, 31: 201–7.
- [33] Chen P, Li W, Yao S, et al. Recognition and counting of wheat mites in wheat fields by a three-step deep learning method. *Neurocomputing*, 2021, 437: 21–30.
- [34] Jiang Q, Wang H, Wang H. Severity assessment of wheat stripe rust based on machine learning. *Front Plant Sci*, 2023, 14: 1150855.
- [35] Li C, Chen S, Ma Y, et al. Wheat pest identification based on deep learning techniques[C]. 2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI), 2024: 87–91.
- [36] 梁佳慧, 蒋倩, 王红丽, 等. 基于机器学习的不同信源小麦叶部病害单病斑图像识别. *植物病理学报*, 2023, 53: 905–21.
- [37] Liang J, Jiang Q, Wang H, et al. Single lesion image recognition of wheat foliar diseases from different information sources based on machine learning. *Acta Phytopathol Sin*, 2023, 53: 905–21.
- [38] Odilbekov F, Armoniene R, Henriksson T, et al. Proximal phenotyping and machine learning methods to identify *Septoria tritici* blotch disease symptoms in wheat. *Front Plant Sci*, 2018, 9: 685.
- [39] Li L, Dong Y, Xiao Y, et al. Combining disease mechanism and machine learning to predict wheat *Fusarium head blight*. *Remote Sens*, 2022, 14: 2732.
- [40] Mustafa G, Zheng H, Li W, et al. *Fusarium head blight* monitoring in wheat ears using machine learning and multimodal data from asymptomatic to symptomatic periods. *Front Plant Sci*, 2022, 13: 1102341.
- [41] Feng ZH, Wang LY, Yang ZQ, et al. Hyperspectral monitoring of powdery mildew disease severity in wheat based on machine learning. *Front Plant Sci*, 2022, 13: 828454.
- [42] Hu YY, Lu HF, Liu WX, et al. Effects of high temperature and water deficiency during grain filling on activities of key starch synthesis enzymes and starch accumulation in wheat. *Acta Agron Sin*, 2018, 44: 591–600.
- [43] Correia PMP, Cairo Westergaard J, Bernardes da Silva A, et al. High-throughput phenotyping of physiological traits for wheat resilience to high temperature and drought stress. *J Exp Bot*, 2022, 73: 5235–51.
- [44] Gupta A, Kaur L, Kaur G. Drought stress detection technique for wheat crop using machine learning. *PeerJ Comput Sci*, 2023, 9: e1268.
- [45] Wen G, Ma BL. Decoding canola and oat crop health and productivity under drought and heat stress using bioelectrical signals and machine learning. *Artif Intell Agric*, 2025, 15: 696–706.
- [46] Liang X, Yu S, Ju Y, et al. Multi-scale remote-sensing phenomics integrated with multi-omics: Advances in crop drought-heat stress tolerance mechanisms and perspectives for climate-smart agriculture. *Plants (Basel)*, 2025, 14: 2829.
- [47] Li M, Wang Z, Cheng T, et al. Phenotyping terminal heat stress tolerance in wheat using UAV multispectral imagery to monitor leaf stay-greenness. *Smart Agric Technol*, 2025, 11: 100996.
- [48] Yang N, Zhang Z, Ding B, et al. Evaluation of winter-wheat water stress with UAV-based multispectral data and ensemble learning method. *Plant Soil*, 2023, 497: 647–68.
- [49] Chandel NS, Rajwade YA, Dubey K, et al. Water stress

- identification of winter wheat crop with state-of-the-art AI techniques and high-resolution thermal-RGB imagery. *Plants (Basel)*, 2022, 11: 3344.
- [49] Zhao H, Wang J, Guo J, et al. Detecting water stress in winter wheat based on multifeature fusion from UAV remote sensing and stacking ensemble learning method. *Remote Sens*, 2024, 16: 4100.
- [50] Qi G, Zhao G, Xi X. Soil salinity inversion of winter wheat areas based on satellite-unmanned aerial vehicle-ground collaborative system in coastal of the yellow river delta. *Sensors (Basel)*, 2020, 20: 6521.
- [51] Fu X, Han B, Liu S, et al. WSVAS: A YOLOv4-based phenotyping platform for automatically detecting the salt tolerance of wheat based on seed germination vigour. *Front Plant Sci*, 2022, 13: 1074360.
- [52] Kecoglu I, Sirkeci M, Unlu MB, et al. Quantification of salt stress in wheat leaves by Raman spectroscopy and machine learning. *Sci Rep*, 2022, 12: 7197.
- [53] 王慧, 曾路生, 孙永红, 等. 重金属铜和锌胁迫下的小麦冠层反射光谱特征. *农业工程学报*, 2017, 33: 171–6. Wang H, Zeng L, Sun Y, et al. Wheat canopy reflectance spectral traits under copper and zinc heavy metal stress. *Transact Chin Soc Agric Engin*, 2017, 33: 171–6.
- [54] Zhai Y, Zhou L, Qi H, et al. Application of visible/near-infrared spectroscopy and hyperspectral imaging with machine learning for high-throughput plant heavy metal stress phenotyping: A review. *Plant Phenomics*, 2023, 5: 0124.
- [55] 牛硕, 李艳玲, 杨阳, 等. 基于机器学习方法的小麦镉富集因子预测. *环境科学*, 2023, 44: 3619–26. Niu S, Li Y, Yang Y, et al. Prediction of wheat cadmium bioaccumulation factors based on machine learning methods. *Environ Sci*, 2023, 44: 3619–26.
- [56] 钟亮, 钱家炜, 储学远, 等. 利用高光谱遥感技术监测小麦土壤重金属污染. *农业工程学报*, 2023, 39: 265–70. Zhong L, Qian J, Chu X, et al. Monitoring of soil heavy metal pollution in wheat fields using hyperspectral remote sensing technology. *Transact Chin Soc Agric Engin*, 2023, 39: 265–70.
- [57] Bali N, Singla A. Deep learning based wheat crop yield prediction model in Punjab Region of North India. *Appl Artif Intell*, 2021, 35: 1304–28.
- [58] 魏萌秀. 基于机器学习的运城市小麦产量预测[D]. 晋中: 山西农业大学, 2022. Wei M. Prediction of wheat yield in Yuncheng city based on machine learning[D]. Jinzhong: Shanxi Agricultural University, 2022.
- [59] 江斌超. 基于机器学习和时间序列的黄淮海小麦产量预测方法研究[D]. 郑州: 河南农业大学, 2022. Jiang B. A study on wheat yield prediction methods in the Huang-Huai-Hai region based on machine learning and time series analysis[D]. Zhengzhou: Henan Agricultural University, 2022.
- [60] 王垒. 基于机器学习与智能优化算法的河南省小麦产量预测方法研究[D]. 郑州: 河南农业大学, 2023. Wang L. A study on wheat yield prediction methods in Henan province based on machine learning and intelligent optimization algorithms[D]. Zhengzhou: Henan Agricultural University, 2023.
- [61] Misra T, Arora A, Marwaha S, et al. Yield-SpikeSegNet: An extension of SpikeSegNet deep-learning approach for the yield estimation in the wheat using visual images. *Appl Artif Intell*, 2022, 36: e2137642.
- [62] Fei S, Hassan MA, Xiao Y, et al. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precis Agric*, 2023, 24: 187–212.
- [63] 齐浩, 吕亮杰, 孙海芳, 等. 基于无人机高光谱与机器学习的小麦品系产量估测研究. *农业机械学报*, 2024, 55: 260–9. Qi H, Lv L, Sun H, et al. Study on yield estimation of wheat lines based on UAV hyperspectral technology and machine learning. *Transact Chin Soc Agric Machinery*, 2024, 55: 260–9.
- [64] Yang S, Li L, Fei S, et al. Wheat yield prediction using machine learning method based on UAV remote sensing data. *Drones*, 2024, 8: 284.
- [65] 张磊. 基于机器学习的多指标小麦品质预测方法研究[D]. 郑州: 河南工业大学, 2021. Zhang L. A study on multi-index wheat quality prediction methods based on machine learning. Zhengzhou: Henan University of Technology, 2021.
- [66] Zhang S, Qi X, Gao M, et al. Estimation of wheat protein content and wet gluten content based on fusion of hyperspectral and RGB sensors using machine learning algorithms. *Food Chem*, 2024, 448: 139103.
- [67] Shi T, Gao Y, Song J, et al. Using VIS-NIR hyperspectral imaging and deep learning for non-destructive high-throughput quantification and visualization of nutrients in wheat grains. *Food Chem*, 2024, 461: 140651.
- [68] 宋茂兴, 马宏亮, 吴志会, 等. 基于SWG和LSG堆叠模型机器学习算法的小麦种子蛋白质含量检测的研究. *分析测试学报*, 2025, 44: 2087–94. Song M, Ma H, Wu Z, et al. Study on detection of protein content in wheat seeds based on SWG and LSG stacking machine learning model. *J Instrument Anal*, 2025, 44: 2087–94.

- [69] 谢上. 基于直接RNA测序和机器学习解码小麦表观转录组[D]. 杨凌: 西北农林科技大学, 2022.
Xie S. Decoding the wheat epitranscriptome based on direct RNA sequencing and machine learning[D]. Yangling: Northwest A&F University, 2022.
- [70] N'Diaye A, Byrns B, Cory AT, et al. Machine learning analyses of methylation profiles uncovers tissue-specific gene expression patterns in wheat. *Plant Genome*, 2020, 13: e20027.
- [71] 吕国豪. 基于机器学习的小麦特征基因筛选和基因组预测研究[D]. 合肥: 安徽农业大学, 2024.
Lv G. Study on wheat candidate gene screening and genomic prediction based on machine learning[D]. Hefei: Anhui Agricultural University, 2024.
- [72] Gonzalez-Camacho JM, Ornella L, Perez-Rodriguez P, et al. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome*, 2018, 11: 170104.
- [73] Sandhu KS, Aoun M, Morris CF, et al. Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. *Biology (Basel)*, 2021, 10: 3835.
- [74] Ibrahim MS, Ibrahim SM. Identifying biotic stress-associated molecular markers in wheat using differential gene expression and machine learning techniques. *Highlight BioSci*, 2022,10: 36462.
- [75] Zhu C, Zhao L, Zhao S, et al. Utilizing machine learning and bioinformatics analysis to identify drought-responsive genes affecting yield in foxtail millet. *Int J Biol Macromol*, 2024, 277: 134288.
- [76] Liu S, Li K, Dai X, et al. A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nat Genet*, 2025, 57: 1008–20.
- [77] Zidi FA, Ouafi A, Bougourzi F, et al. Advancing wheat crop analysis: A survey of deep learning approaches using hyperspectral imaging. *Comput Electron Agric*, 2025, 238: 110770.
- [78] Barbedo JGA. A review on the combination of deep learning techniques with proximal hyperspectral images in agriculture. *Comput Electron Agric*, 2023, 210: 107920.
- [79] Barbedo JGA. A review of artificial intelligence techniques for wheat crop monitoring and management. *Agronomy*, 2025, 15: 1157.
- [80] Li J, Kang J, Lu J, et al. Dynamic gating-enhanced deep learning model with multi-source remote sensing synergy for optimizing wheat yield estimation. *Front Plant Sci*, 2025, 16: 1640806.
- [81] Li Y, Che Y, Zhang H, et al. Wheat growth stage identification method based on multimodal data. *Eur J Agron*, 2025, 162: 127423.
- [82] Yang G, Li Y, He Y, et al. Multimodal large language model for wheat breeding: A new exploration of smart breeding. *ISPRS J Photogramm Remote Sens*, 2025, 225: 492–513.
- [83] Guo F, Wang P, Tansey K, et al. A novel transformer-based neural network under model interpretability for improving wheat yield estimation using remotely sensed multi-variables. *Comput Electron Agric*, 2024, 223: 109111.
- [84] Dong M, Liu S, Jiang R, et al. Comparing and combining data-driven and model-driven approaches to monitor wheat green area index with high spatio-temporal resolution satellites. *Remote Sens Environ*, 2024, 305: 114118.
- [85] 江敏, 金之庆, 葛道阔, 等. CERES-Wheat 模型在我国冬小麦主产区的适用性验证及订正. *江苏农学院学报*, 1998, 19: 64–7.
Jiang M, Jin Z, Ge D, et al. Validation and calibration of CERES-wheat model in major winter wheat production regions of China. *J Jiangsu Agric Coll*, 1998, 19: 64–7.
- [86] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. *PMLR*, 2017: 1126–35.
- [87] Yu T, Zhang H, Chen S, et al. EXGEP: a framework for predicting genotype-by-environment interactions using ensembles of explainable machine-learning models. *Brief Bioinform*, 2025, 26: bbaf414.
- [88] 王鹏新, 王颖, 田惠仁, 等. 基于LightGBM的冬小麦产量估测与可解释性研究. *农业机械学报*, 2023, 54: 197–206.
Wang P, Wang Y, Tian H, et al. Winter wheat yield estimation and interpretability research based on LightGBM. *Transact Chin Soc Agric Machinery*, 2023, 54: 197–206.
- [89] Ryo M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artif Intell Agric*, 2022, 6: 257–65.
- [90] Qiu Z, Wang F, Wang W, et al. YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model. *Front Plant Sci*, 2024, 15: 1495222.
- [91] Qu Z, Liang X, Liang S, et al. Star-YOLO: A lightweight real-time wheat grain detection model for embedded deployment. *IEEE Access*, 2025, 13: 130147–58.

-
- [92] Wang Y, Li C, Jiang C, et al. CropCLR-wheat: A label-efficient contrastive learning architecture for lightweight wheat pest detection. *Insects*, 2025, 16: 1096.
- [93] Khaki S, Safaei N, Pham H, et al. WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing*, 2022, 489: 78–89.
- [94] Dey T, Bera S, Paul B, et al. Fly: Femtolet-based edge-cloud framework for crop yield prediction using bidirectional long short-term memory. *Software Pract Exper*, 2024, 54: 1361–77.
- [95] Wang C, Zhang D, Ma Y, et al. WheatGP, a genomic prediction method based on CNN and LSTM. *Brief Bioinform*, 2025, 26: bbaf191.