

网络出版时间:2025-09-24

网络出版地址:https://link.cnki.net/urlid/61.1359.s.20250923.1414.014

基于新型极限堆叠泛化算法的小麦育种田产量预测研究

姚东泽¹, 费帅鹏¹, 李雷¹, 贾艺丹¹, 王多霞¹, 韩桐鹤^{1,2}, 张博涵^{1,2}, 杨梦娇^{1,3}, 肖永贵¹

(1. 中国农业科学院作物科学研究所/国家小麦改良中心/中国作物基因资源与育种全国重点实验室, 北京 100081;

2. 甘肃农业大学农学院, 甘肃兰州 730070; 3. 新疆农业职业技术大学, 新疆昌吉 831100)

摘要:针对不同小麦品种产量预测中光谱特征差异利用不足和传统集成学习方法泛化能力有限的问题,本研究提出一种新型极限堆叠泛化算法(extreme stacked generalization, ESG),通过动态优化特征选择与模型组合解析品种间微米级光谱差异,以提高在多品种、跨年份数据中的预测精度和稳定性。基于2018—2020年两个小麦生长季的冠层高光谱反射率数据(覆盖灌浆前期和中期)和产量数据,采用跨年份独立验证(框架1)与多时期融合验证(框架2)双验证体系,以平均绝对误差(mean absolute error, MAE)和均方根误差(root mean squared error, RMSE)评估岭回归(ridge regression, RR)、K最邻近回归(K-nearest neighbors, KNN)、随机森林(random forest, RF)、堆叠泛化算法(stacked generalization, SG)和ESG算法性能。结果表明,ESG算法显著优于其他算法,其预测RMSE降至 $1.01 \sim 1.31 \text{ t} \cdot \text{hm}^{-2}$;跨年份预测的稳定性显著提升,误差波动幅度缩减18.9%,显示出较强的环境适应性和品种区分能力;灌浆中期为最优预测窗口期(RMSE = $1.01 \text{ t} \cdot \text{hm}^{-2}$)。这说明ESG算法能够有效适应不同小麦品种的特异性光谱特征和环境变异,实现跨品种、跨年度的产量稳定预测。

关键词: 小麦; 育种田; 产量预测; 极限堆叠泛化算法

中图分类号: S512.1; S314

文献标识码: A

文章编号: 1009-1041(2026)04-0531-10

Yield Prediction of Wheat Breeding Plots Based on a Novel Extreme Stacked Generalization Algorithm

YAO Dongze¹, FEI Shuaipeng¹, LI Lei¹, JIA Yidan¹, WANG Duoxia¹, HAN Tonghe^{1,2},
ZHANG Bohan^{1,2}, YANG Mengjiao^{1,3}, XIAO Yonggui¹

(1. State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, National Wheat Improvement Centre, Chinese Academy of Agricultural Sciences(CAAS), Beijing, 100081, China; 2. College of Agronomy, Gansu Agricultural University, Lanzhou, Gansu 730070, China; 3. Xinjiang Vocational University of Agriculture, Changji, Xinjiang 831100, China)

Abstract: Addressing the challenges of inadequate utilization of spectral feature variations among wheat cultivars and the limited generalization capacity of conventional ensemble learning methods for yield prediction, this study proposed a novel Extreme Stacked Generalization(ESG) algorithm. The ESG method was designed to enhance prediction accuracy and stability across multiple cultivars and years by dynamically optimizing feature selection and model integration, thereby resolving micrometer-level spectral discrepancies among cultivars. The study utilized canopy hyperspectral reflectance and yield data collected over two wheat growing seasons(2018–2020), containing the early and middle grain-filling stages. A dual-validation framework was employed, comprising independent cross-year validation(Framework 1) and multi-period data fusion validation(Framework 2). The performance of the

收稿日期: 2025-06-09

修回日期: 2025-07-08

基金项目: 科技创新 2030—“新一代人工智能”重大项目(2022ZD0115703); 国家自然科学基金项目(32372196); 科技创新工程—高产高效技术集成与示范应用项目(Y2025WC03); 庭州青年科技创新人才项目(2023QN01)

第一作者 E-mail: 15500351725@163.com(姚东泽)

通讯作者 E-mail: yangmengjiao000@163.com(杨梦娇); xiaoyonggui@caas.cn(肖永贵)

ESG algorithm was benchmarked against Ridge Regression(RR), K-Nearest Neighbors(KNN), Random Forest(RF), and a standard Stacked Generalization(SG) algorithm, using Mean Absolute Error (MAE) and Root Mean Squared Error(RMSE) as evaluation metrics. Results demonstrated that the ESG algorithm significantly outperformed all other models, reducing the prediction RMSE to a range of $1.01\text{--}1.31\text{ t}\cdot\text{hm}^{-2}$. The stability of cross-year predictions was notably improved, with the range of error fluctuation decreasing by 18.9%, indicating strong environmental adaptability and cultivar-discriminating capabilities. Furthermore, the middle grain-filling stage was identified as the optimal prediction window, achieving a RMSE of $1.01\text{ t}\cdot\text{hm}^{-2}$. This study concludes that the ESG algorithm can effectively adapt to the specific spectral characteristics of different wheat cultivars and environmental variations, enabling robust and stable yield prediction across diverse cultivars and growing years.

Keywords: Wheat; Breeding plots; Yield Prediction; Extreme stacked generalization algorithm

小麦是中国第三大粮食作物,产量约占全国谷物总产量的20%,在国家粮食安全中占有重要地位^[1]。面对气候变化与极端天气频发的挑战,培育高产稳产品种成为关键^[2]。传统育种依赖人工进行表型鉴定,存在效率低、标准不一、破坏性强等问题,难以满足现代农业需求。随着无人机、田间机器人等平台的发展,搭载光谱设备可实现作物生长的高效监测与产量评估^[3-4],为精准育种提供技术支持^[5-7]。

目前,卫星与无人机高光谱遥感虽能实现大范围监测,但光谱分辨率较低(10~60 nm),难以捕捉小麦品种间细微的生理差异。相比之下,地面高光谱辐射仪(如 ASD FieldSpec)可在 350~2 500 nm 范围内提供连续窄波段反射光谱,灵敏反映叶片生化组分变化^[8-9],结合温度与结构信息,可更准确地评估作物生长状态,且受天气影响较小,数据可靠性高^[10-12]。

基于高光谱数据构建产量预测模型已成为研究热点^[13-14]。支持向量机^[15]、K 近邻^[5-19]、高斯过程^[20]等机器学习算法在处理高维数据方面表现良好^[16-18],但单一模型在应对复杂非线性光谱关系时存在局限,如线性模型难以刻画波段间交互,支持向量机等局部算法对全局趋势敏感性不足。集成学习算法如随机森林和堆叠泛化(stacked generalization, SG)虽可提升模型表达能力^[21-31],但其固定基模型结构难以适应异构、跨年度数据,泛化性能受限。

针对上述问题,本研究以黄淮海麦区主栽小麦品种为对象,采集 2018—2020 年两个小麦生长季灌浆期的冠层全波段高光谱数据,提出一种改进的集成学习算法即极限堆叠泛化(extreme

stacked generalization, ESG)。该算法通过动态穷举模型组合与自适应加权机制,突破传统 SG 的固定结构限制,增强对多源异构数据的融合能力。本文重点探讨了 ESG 中动态特征选择对品种特异性光谱特征的提取效果,以及自适应加权在跨年度预测中的泛化优势,同时通过分析不同时间窗口的预测精度,确定灌浆期最佳光谱采集时期,以期构建稳定、高效的小麦产量预测模型,为精准育种提供技术支撑。

1 材料与方法

1.1 试验材料与田间设计

本研究选用黄淮麦区小麦主栽品种 100 个,2018 年和 2019 年 10 月种植于中国农业科学院作物科学研究所新乡试验基地(113°51'E, 35°18'N)。试验区海拔高度 75 m,年均温度 14.3℃,属于温带大陆季风气候,无霜期 191 d。田间试验设置正常和节水两种灌溉处理,每种处理设置两次重复。节水灌溉处理在小麦分蘖期、拔节期、抽穗期和灌浆期的灌水量均为 250 mm;正常灌溉处理分蘖期的灌水量为 250 mm,拔节期、抽穗期和灌浆期的灌水量均为 500 mm。小区面积为 4.2 m²(3.0 m × 1.4 m),每个小区种植 6 行小麦,行间距为 0.2 m。试验两年均采用完全随机区组设计。出苗后对缺苗断垅处进行移栽,确保苗全苗匀。成熟后,所有小区均于 6 月初采用小区联合收割机收获,每个小区单独装袋、晾晒、脱水,在 12.5%的水分条件下称重测定产量。

1.2 冠层反射率的采集

本研究使用高光谱辐射仪(FieldSpec 3, Analytical Spectral Devices ASD, Boulder, CO, U-

nited States) 采集冠层光谱数据。该设备配备 25° 视场光纤,覆盖范围为 350~2 500 nm,采样间隔为 1.4 nm(350~1 000 nm)和 2 nm(1 000~2 500 nm),重采样间隔为 1 nm,包括 2 151 个波段。为了获得时序一致、生理表征能力强、噪音干扰少的光谱数据^[32-34],选择在小麦灌浆前期和灌浆中期晴朗、无云的天气(11:00 至 13:00),在每个小区选择 4 个代表性位置进行光谱测定。

采集时将传感器探头垂直向下置于作物冠层上方 100 cm 处,每个位置采集 10 次光谱读数,每个小区共采集 40 个读数,求平均值作为小区最终值。为了防止在整个测量过程中由于太阳位置的变化而引起的误差,每隔大约 10 个小区使用漫反射标准白板进行一次辐射校准。

使用 View Spec 软件(ASD Inc. Boulder, CO, United States)提取光谱曲线,并计算多条光谱曲线的平均值,得到输出反射率文件^[35]。利用 Savitzky-Golay 滤波器通过平滑去噪来增强高光谱数据的平滑度,从而增强光谱一致性,减少噪声干扰。此外,去除大气吸水显著的光谱区域(1 350~1 500、1 800~2 100、2 350~2 500 nm)干扰波段^[36],最终保留 350~2 500 nm 范围内 1 980 个有效波段作为输入变量。

1.3 建模方法

1.3.1 机器学习算法

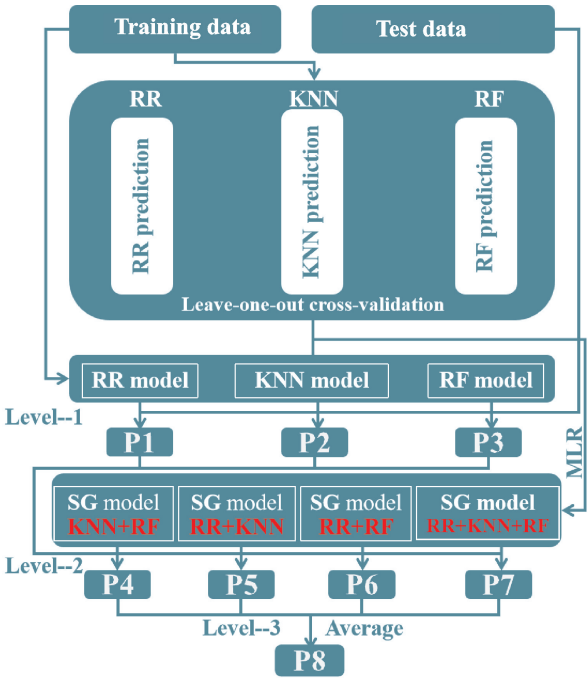
本研究将岭回归(ridge regression, RR)、K 最邻近回归(K-nearest neighbors regression, KNN)和随机森林(random forest, RF)作为基础回归方法,通过 R 语言中的“caret”包实现,其超参数使用网格搜索和 10 倍交叉验证进行微调。每种机器学习方法的超参数候选范围由“train”函数自动设置。这三种常用的机器学习算法中,RR 算法是一种针对共线性问题的线性回归改良算法,特别适用于特征数量多于样本数量的情况^[37];KNN 是一种基于实例的非参数回归算法,其原理是从训练集中选取与预测样本距离最近的 k 个样本,通过加权平均或直接平均这些样本输出值得到预测值^[38];RF 是一种集成学习算法,其通过构建多个决策树并进行整合来提高预测的准确性和鲁棒性,预测结果是所有单独决策树预测结果的平均值^[39]。

1.3.2 极限堆叠泛化算法(ESG)

SG 是一种集成学习算法,已广泛用于解决回归问题^[28-31]。在 SG 的基础上,提出了极限堆

叠泛化算法(ESG),此算法的核心原则是采用 RR、KNN 和 RF 算法作为基础学习器。将所有可能的模型组合由第二层组合到第三层。ESG 的具体步骤如下(图 1):

- (1)分别使用 RF、RR 和 KNN 对训练数据进行留一交叉验证,生成三组样本外预测值。
- (2)将这三种机器学习模型的预测值组合成 KNN+RF、RR+KNN、RR+RF、RR+KNN+RF、4 个 $(C_3^2+C_3^3)$ 预测矩阵。
- (3)训练集上的标签与这 4 个组合在一起形成预测矩阵,并用于训练 4 个不同的多元线性回归(MLR)模型。
- (4)使用完整的训练样本训练 RR、KNN 和 RF 模型,并在测试样本上生成相应的预测值。
- (5)将步骤(3)的预测值作为步骤(2)的 MLR 模型的检验数据,生成四组预测值。然后对这四组预测值进行平均,得到最终的预测结果。
- (6)将 ESG 扩展到一般情况,假设有 M 个基础学习模型,这将产生总共 $C_M^2 + C_M^3 + C_M^4 + C_M^5 + \dots + C_M^{M-1} + C_M^M$ 个组合。最终的预测结果是所有模型组合预测结果的平均值。



RF: 随机森林; KNN: K 最邻近回归; RR: 岭回归; MLR: 多元线性回归; P1~P7 表示不同建模阶段的预测。
RF: Randomforest; KNN: K-nearest neighbors regression; RR: Ridge regression; MLR: Multiple linear regression; P1~P7 represent predictions at different modeling stages

图 1 极限堆叠泛化框架
Fig. 1 Extreme stacked generalization framework

1.4 建模框架和精度评估

为确保产量预测模型的准确性与可靠性,本研究设计了两种模型验证框架,用以评估不同生长阶段小麦产量预测模型的性能。首先,采用跨生长季数据构建验证框架 1,其中验证框架 1-1 和验证框架 1-2 分别以 2019 年的灌浆前期及中期数据作为训练集,对应 2020 年的数据作为测试集,以检验模型对不同年份数据的泛化能力。其次,通过数据融合构建验证框架 2,分别将 2019、2020 年的灌浆前期和中期数据合并,并按照 7:3 的比例划分为训练集和验证集,形成验证框架 2-1 和验证框架 2-2,以验证模型在不同数据条件下的稳定性。两种验证框架具有明确的互补价值,四种验证子框架设计能够全面评估和比较模型在不同数据集和生长条件下的表现,旨在评估模型对特定生长阶段数据的整体泛化性和鲁棒性。

基于此,针对这四个验证框架分别构建冬小麦产量预测回归初级模型,再将初级模型的输出作为下一级运用 SG 算法的输入参数,按照四个预测矩阵($C_3^2 + C_3^3$)对新数据进行二次训练,随后引入 ESG 算法,构建四个验证框架下冬小麦产量预测最终模型。将 ESG 算法在四个建模框架上进行了评估,并与三个基础机器学习模型和标准 SG 方法进行比较。在模型评价中,采用平均绝对误差(mean absolute error, MAE)和均方根误差(root mean squared error, RMSE)评估预测精度。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

其中, n 是样本数量, y_i 是第 i 个实际值, $\hat{y}_i$ 是第 i 个预测值。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

其中, n 是样本数量, y_i 是第 i 个实际值, $\hat{y}_i$ 是第 i 个预测值。

2 结果与分析

2.1 小麦产量分布特征

两个小麦生长季各涵盖 400 个样本,产量均呈近似正态分布(图 2)。2019 和 2020 年平均产量分别为 8.71 和 $7.19 \text{ t} \cdot \text{hm}^{-2}$,变化范围分别为 $5.0 \sim 12.0$ 和 $5.0 \sim 10.5 \text{ t} \cdot \text{hm}^{-2}$,变异系数分别为 17.92% 和 16.36% ,两年平均产量存在显著差异,变异系数接近。2019 年水温条件比较适合小麦光合作用,有利于产量形成,而 2020 年出

现极端干旱事件,因此两年的产量差异应主要归因于年度间气候因素的影响。

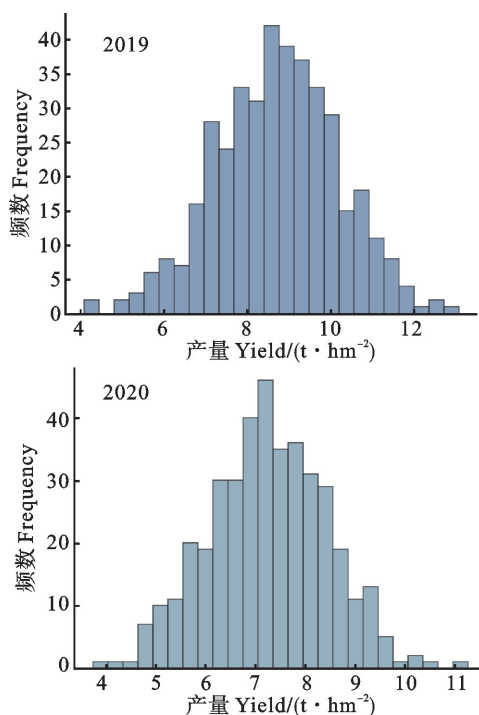


图 2 2019 和 2020 年小麦产量分布

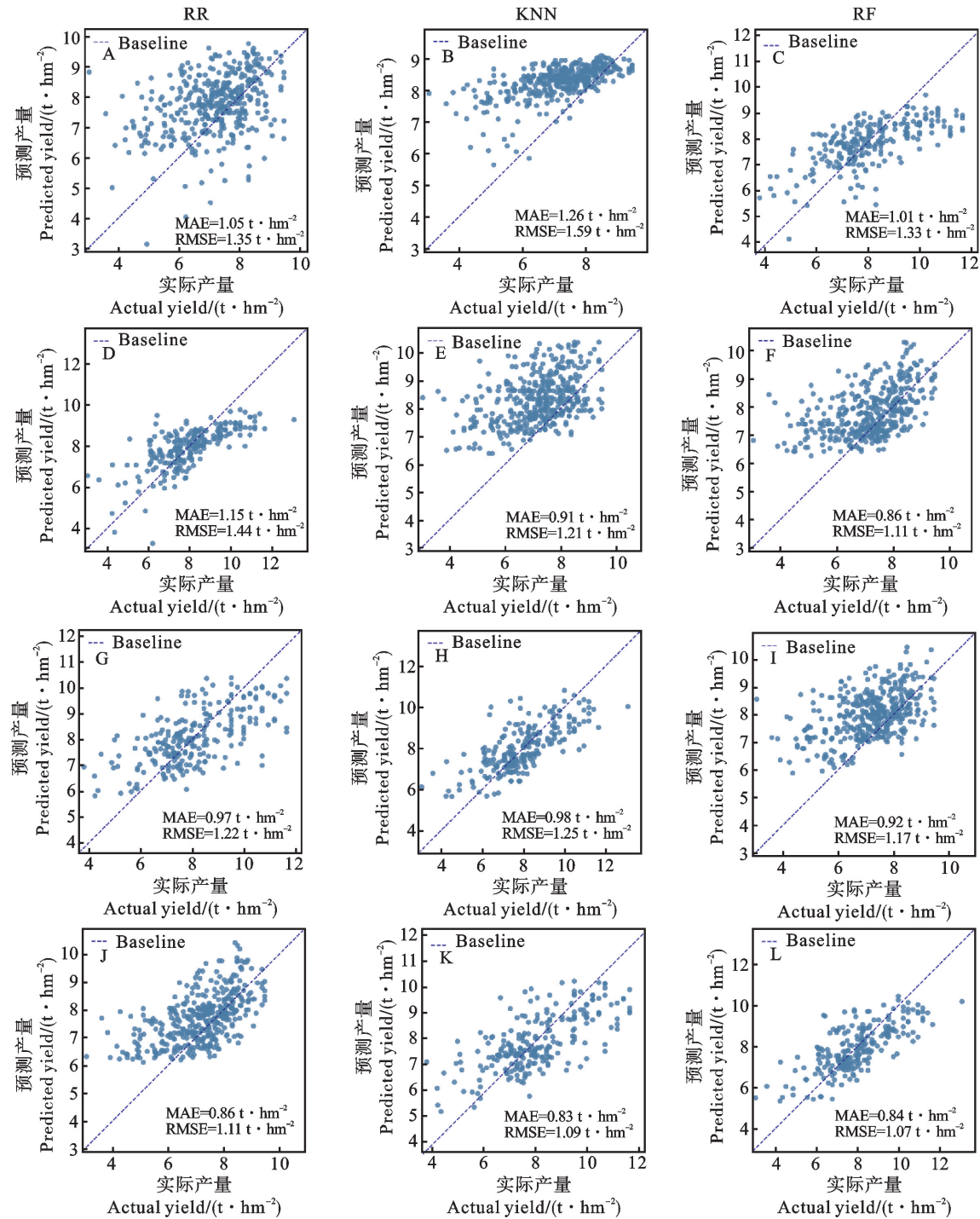
Fig. 2 Distribution of wheat yield in 2019 and 2020

2.2 基础模型性能评估

基于 RR、KNN、RF 机器学习算法分别构建冬小麦产量初级预测模型。从各模型的预测结果(图 3 和表 1)看,RF 在多数框架中表现最优,但其预测偏差在不同产量区间呈现显著差异。例如,在验证框架 1-1 中,RF 模型的预测值在低产区间($<7 \text{ t} \cdot \text{hm}^{-2}$)出现高估(图 3C),数据点密集分布于 1:1 线以上;在验证框架 1-2 中,RR 模型预测精度最低,MAE 和 RMSE 分别为 1.15 和 $1.44 \text{ t} \cdot \text{hm}^{-2}$,而 RF 模型的预测点沿 1:1 线均匀分布(图 3F),MAE 和 RMSE 分别为 0.86 和 $1.11 \text{ t} \cdot \text{hm}^{-2}$;在验证框架 2-2 中,KNN 模型的 MAE 虽然在三种模型中最低($0.83 \text{ t} \cdot \text{hm}^{-2}$),但其预测值在中等产量区间($7 \sim 9 \text{ t} \cdot \text{hm}^{-2}$)的离散度较高(图 3K),说明该算法对全局趋势的敏感性不足。整体上看,RF 模型表现出较好的预测精度。

2.3 标准堆叠泛化算法(SG)性能评估

在三个基础模型的算法上引入标准堆叠泛化(SG)算法后,模型精度进一步提升,但其性能受限于固定组合策略。在验证框架 1-2(灌浆中期)中,RR+KNN+RF 组合的 MAE 和 RMSE 分别为 0.80 和 $1.02 \text{ t} \cdot \text{hm}^{-2}$ (表 2),数据点紧密围绕



A~C: 验证框架 1-1; D~F: 验证框架 1-2; G~I: 验证框架 2-1; J~L: 验证框架 2-2。
A~C: Validation framework 1-1; D~F: Validation framework 1-2; G~I: Validation framework 2-1; J~L: Validation framework 2-2.

图 3 不同模型对小麦产量的预测散点图

Fig. 3 Scatter plot of different models predicting wheat yield

1:1 线分布(图 4H),显著优于单一模型。然而,在跨年份验证框架 1-1 中,SG 算法的 RMSE 波动范围达 1.27~1.39 t · hm⁻²,表明其对年度气候变异的适应性有限。此外,验证框架 2-1 中 RR + KNN 组合虽在多数区间表现稳定,但在高产

区间(>10 t · hm⁻²)数据点下偏趋势,对小麦产量仍存在显著低估现象(图 4J)。

2.4 极限堆叠泛化(ESG)性能评估

ESG 算法通过多模型协同与自适应加权,动态融合所有模型子集(图 5),进一步优化了模型

表 1 基于三种机器学习算法的模型预测精度

Table 1 Prediction accuracy of different models with three machine learning algorithms

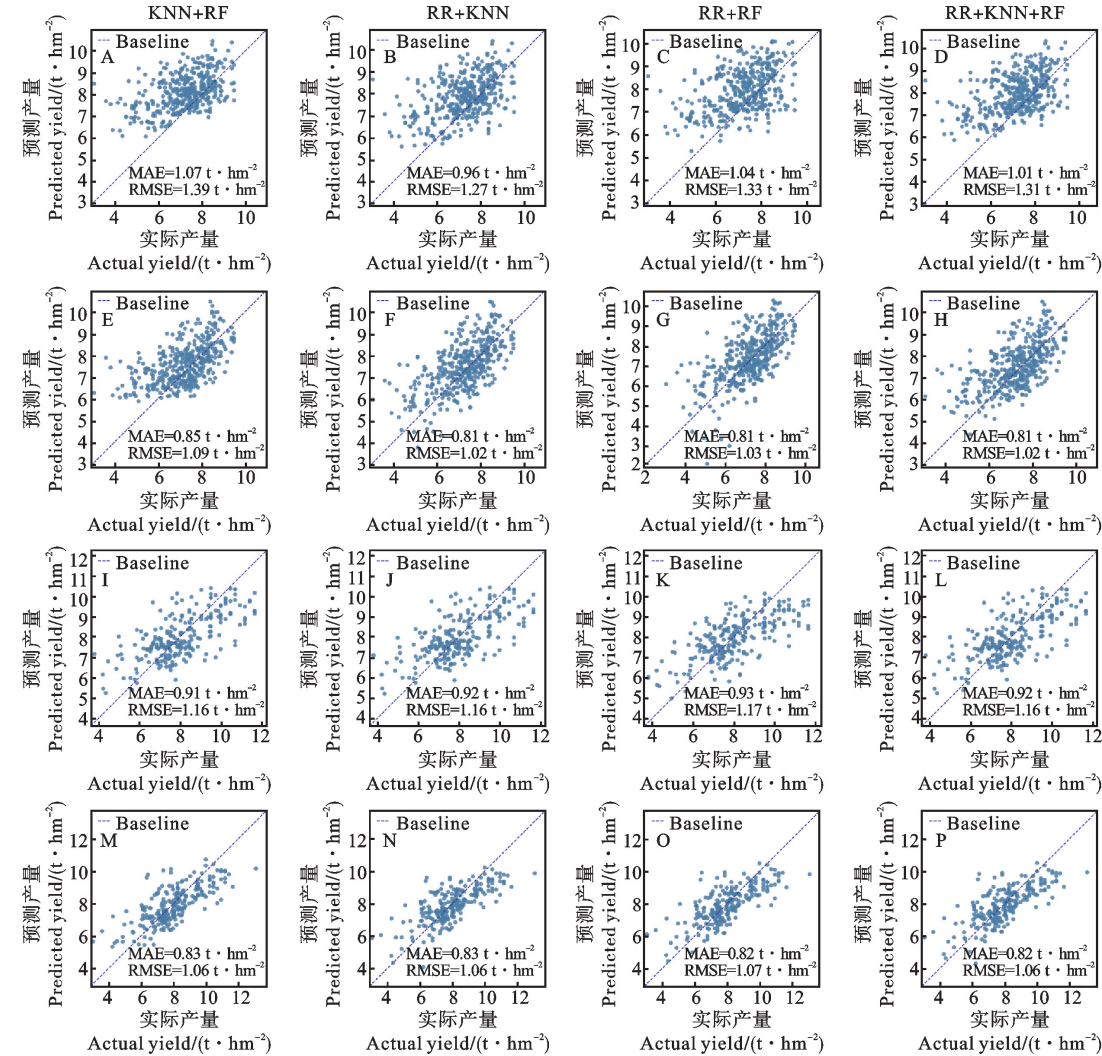
验证框架 Validation framework	模型 Model	评价系数 Evaluation metric	
		MAE/(t · hm ⁻²)	RMSE/(t · hm ⁻²)
1-1	RR	1.05	1.35
	KNN	1.26	1.59
	RF	1.01	1.33
1-2	RR	1.15	1.44
	KNN	0.91	1.21
	RF	0.86	1.11
2-1	RR	0.97	1.22
	KNN	0.98	1.25
	RF	0.92	1.17
2-2	RR	0.86	1.11
	KNN	0.83	1.09
	RF	0.84	1.07

RF:随机森林;KNN:K最近回归;RR:岭回归。下表图同。
RF: Random forest; KNN: K-nearest neighbors; RR: Ridge regression. The same in tables 2-4 and figures 3-6.

表 2 基于 SG 算法的模型预测精度

Table 2 Model prediction accuracy based on SG algorithm

验证框架 Validation framework	模型 Model	评价系数 Evaluation metric	
		MAE/(t · hm ⁻²)	RMSE/(t · hm ⁻²)
1-1	KNN+RF	1.07	1.39
	RR+KNN	0.96	1.27
	RR+RF	1.04	1.33
	RR+KNN+RF	1.01	1.31
1-2	KNN+RF	0.85	1.09
	RR+KNN	0.81	1.02
	RR+RF	0.81	1.03
	RR+KNN+RF	0.80	1.02
2-1	KNN+RF	0.92	1.16
	RR+KNN	0.91	1.16
	RR+RF	0.93	1.17
	RR+KNN+RF	0.92	1.16
2-2	KNN+RF	0.83	1.06
	RR+KNN	0.83	1.06
	RR+RF	0.82	1.07
	RR+KNN+RF	0.82	1.06



A~D:验证框架 1-1;E~H:验证框架 1-2;I~L:验证框架 2-1;M~P:验证框架 2-2。

A~D: Validation framework 1-1; E~H: Validation framework 1-2; I~L: Validation framework 2-1; M~P: Validation framework 2-2.

图 4 SG 模型对小麦产量预测的散点图
Fig. 4 Scatter plot of wheat yield prediction based on SG model

预测性能的稳定性。例如,在验证框架 1-2 中,ESG 的 RMSE 为 $1.01 \text{ t} \cdot \text{hm}^{-2}$,虽然较 SG 最优组合(RR+KNN+RF)仅降低 1.0%,但其跨年份误差波动范围($\Delta\text{RMSE}=0.30 \text{ t} \cdot \text{hm}^{-2}$)显著低于 SG($0.37 \text{ t} \cdot \text{hm}^{-2}$)(表 3),稳定性相对提升 18.9%。其预测散点图中数据点分布更集中,尤其是在高产区间($>9 \text{ t} \cdot \text{hm}^{-2}$)的偏离程度减少。此外,灌浆中期的(RMSE= $1.01 \text{ t} \cdot \text{hm}^{-2}$)较灌浆前期降低 22.9%。尽管 ESG 在绝对误差降幅上有限,但其通过增强模型多样性,显著缓解了传统建模方法对异构数据的适应性问题。

2.5 综合性能总结与关键窗口期验证

从各层级模型的整合结果(图 6)看,融合验证的预测精度显著优于跨生长季独立验证,框架 2 中 ESG 的 RMSE 范围($1.06 \sim 1.15 \text{ t} \cdot \text{hm}^{-2}$)较框架 1 整体降低 30%,表明跨年份泛化能力是模型优化的核心方向。同时,灌浆中期数据的预测误差在所有框架中均最低,在散点图中灌浆中期数据点的均匀分布(图 3F、图 5B)与前期数据的系统性偏差(图 3C、图 5A)形成鲜明对比,印证了前人关于灌浆期光谱与产量关联性的理论,也为育种资源的高效配置提供了实践依据。

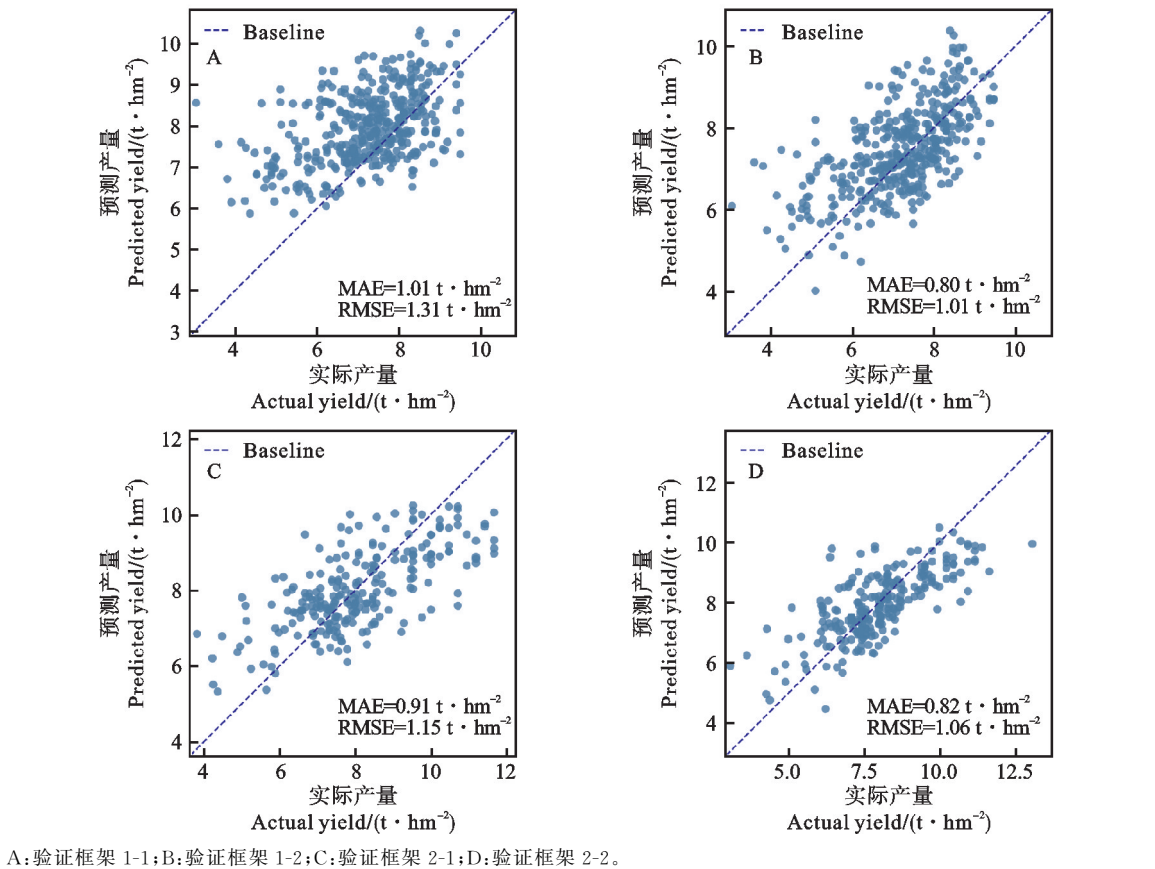
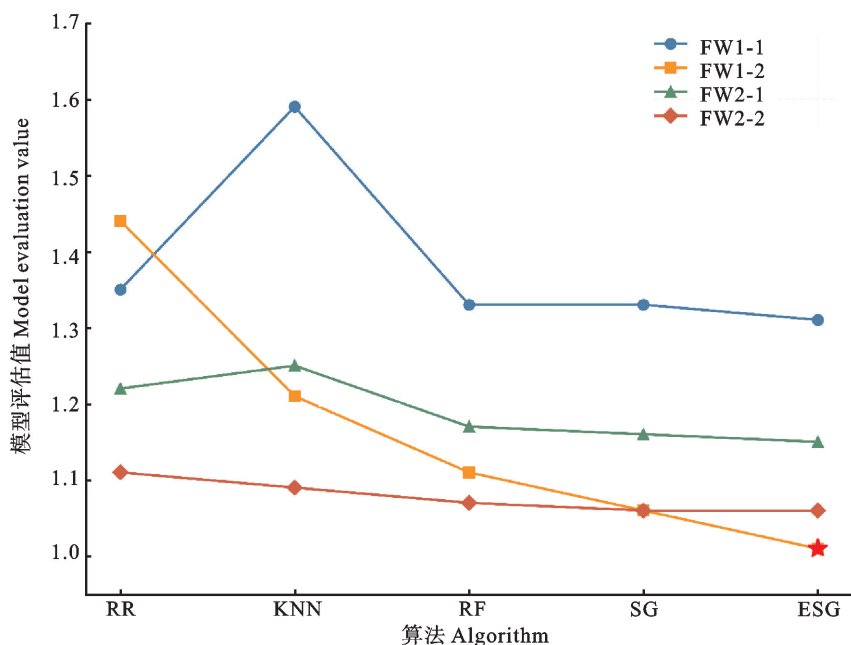


图 5 ESG 模型对小麦产量预测的散点图

Fig. 5 Scatter plot of wheat yield prediction based on ESG model

表 3 基于 ESG 算法的模型预测精度

验证框架 Validation framework	模型 Model	评价系数 Evaluation metric	
		MAE/($\text{t} \cdot \text{hm}^{-2}$)	RMSE/($\text{t} \cdot \text{hm}^{-2}$)
1-1	(SG1+SG2+SG3+SG4)/4	1.01	1.31
1-2	(SG1+SG2+SG3+SG4)/4	0.80	1.01
2-1	(SG1+SG2+SG3+SG4)/4	0.91	1.15
2-2	(SG1+SG2+SG3+SG4)/4	0.82	1.06



FW1-1;验证框架 1-1;FW1-2;验证框架 1-2;FW2-1;验证框架 2-1;FW2-2;验证框架 2-2。

FW1-1; Validation framework 1-1; FW1-2; Validation framework 1-2; FW2-1; Validation framework 2-1; FW2-2; Validation framework 2-2.

图6 不同算法模型在四个验证框架下的 RMSE

Fig. 6 Model RMSE based on ESG algorithm under four validation frameworks

3 讨论

3.1 生育时期选择与模型构建的科学依据

小麦产量的精确预测对于提高育种效率和加快育种进程至关重要。在作物不同的生长阶段,地面空间异质性会发生变化,包括光照、温度、水分和养分等,这种变化对作物的冠层光谱指数与产量之间的相关性具有显著影响。本研究提出了“灌浆前期和中期双阶段融合策略”,解决了单一生育时期数据代表性不足的问题。选择灌浆前期与中期作为核心数据采集窗口,其科学依据源于作物生理特性与光谱响应的动态关联性。前人研究表明,在不同生长阶段作物的光谱指数与产量的相关性也会有所不同。灌浆期是小麦产量形成的关键阶段,此时期冠层光谱特征与籽粒干物质积累显著相关^[40]。尽管已有文献多聚焦于开花期至灌浆期的对比^[41],但本试验中 100 个品种的生育时期差异显著,抽穗至灌浆时间跨度达 15 d,导致开花期数据因品种间物候异步性难以统一建模。灌浆前期与中期数据可有效规避此干扰,因其对应籽粒快速充实的稳定阶段,光谱特征与产量的时空一致性较高。其策略与 Maimaitijiang 等^[10]提出的“关键窗口期”理论一致,即选择生理过程稳定的时期可最大化光谱信息的预测效力。

3.2 ESG 算法创新与光谱数据协同效应

本研究提出的极限堆叠泛化算法(ESG)通过动态穷举所有可能的模型组合并自适应优化权重,突破了传统堆叠泛化(SG)固定架构的局限性,为解决高光谱数据非平稳性问题提供了创新性解决方案。ESG 算法的核心优势体现在其能够充分利用异构基模型(RR、RF、KNN)的互补特性,其中 RR 擅长全局线性拟合,RF 专注于非线性特征交互,而 KNN 则能捕捉局部响应模式,其协同作用使得模型对高光谱数据的解析能力显著提升。与单一机器学习模型相比,ESG 在 MAE 和 RMSE 上分别实现了 6.2%~20.6%和 5.7%~17.6%的降低(表 2、表 4),特别是在跨年度验证中表现出较强的稳定性,误差波动范围(Δ RMSE)较 SG 缩减了 18.9%,其结果与 Liang 等^[42-43]在基因组预测中的研究发现相一致。

ESG 算法的动态组合机制使其对环境变异(如干旱、温度波动)具有较强的鲁棒性,其特性与 Feng 等^[44]在多模型融合研究中观察到的现象一致。虽然相较于卫星平台的多源融合方法,ESG 在单一高光谱数据下的绝对精度($\text{RMSE}=1.01 \text{ t} \cdot \text{hm}^{-2}$)略低,但其在数据维度和计算效率上的优势(数据需求减少 30%)使其更适用于育种田的精细化需求。其中,灌浆中期的预测性能最优

($RMSE=1.01\text{ t}\cdot\text{hm}^{-2}$), 与该时期小麦生理过程与光谱特征的稳定对应关系密切相关。

3.3 研究局限与未来方向

本研究采用的双验证框架(跨年度独立验证与多时期融合验证)为评估模型性能提供了全面视角。其中, 框架1通过严格的跨年测试证实了模型对气候变异的适应能力, 而框架2则验证了模型在不同数据条件下的稳定性。这种双重验证机制不仅确保了结论的可靠性, 也为育种实践中的模型应用提供了重要参考。尽管ESG算法在小麦产量预测中表现出色, 但其应用潜力不仅限于此。近期研究显示, 类似方法在玉米、水稻等其他作物的表型分析中也取得了成功, 表明ESG的架构设计具有跨作物的普适性^[45-46]。

当前研究仍存在一些局限性, 如高光谱数据采集易受环境干扰、算法计算复杂度较高等, 这将是未来研究工作的重点方向。通过引入实时辐射校准、开发自适应组合筛选策略, 以及融合激光雷达等新型传感数据, 有望进一步提升模型的预测性能和适用性。未来研究可从3个方向进一步拓展: 首先, 引入更多作物类型的公共数据集验证算法的迁移能力; 其次, 探索与无人机多光谱或卫星时序数据的融合策略^[47-48], 构建“光谱-结构”双驱动模型; 最后, 优化算法效率以应对更大规模的数据处理需求。这些改进将有助于建立更加通用、高效的作物表型分析框架。

4 结论

本研究通过ESG集成算法创新与高光谱数据的协同建模, 实现了对不同小麦品种产量预测精度与跨生长季稳定预测的双重提升。其中, 灌浆中期冠层光谱数据预测误差($RMSE=1.01\text{ t}\cdot\text{hm}^{-2}$)较灌浆前期($RMSE=1.31\text{ t}\cdot\text{hm}^{-2}$)降低22.9%, 说明灌浆前期为产量预测的最优窗口期。ESG算法通过动态穷举基模型组合, 在跨年度独立验证(框架1)中将误差波动范围从SG的 $0.37\text{ t}\cdot\text{hm}^{-2}$ 压缩至 $0.30\text{ t}\cdot\text{hm}^{-2}$ (提升18.9%), 在数据融合框架(框架2)中从 0.10 降至 $0.09\text{ t}\cdot\text{hm}^{-2}$ (提升10%), 有效缓解了固定模型组合对数据非平稳性的敏感性问题。

参考文献:

- [1] 魏益民, 张波, 关二旗, 等. 中国冬小麦品质改良研究进展[J]. 中国农业科学, 2013, 46(20): 4189.
- [2] AWASTHI R, KAUSHAL N, VADEZ V, et al. Individual and combined effects of transient drought and heat stress on carbon assimilation and seed filling in chickpea [J]. *Functional Plant Biology*, 2014, 41(11): 1148.
- [3] LI Q, JIN S C, ZANG J R, et al. Deciphering the contributions of spectral and structural data to wheat yield estimation from proximal sensing [J]. *The Crop Journal*, 2022, 10(5): 1334.
- [4] WANG T, ZHANG R, FRITSCHI F B, et al. Fusion of Sentinel-2 and PlanetScope data for smallholder crop yield mapping [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 207: 112.
- [5] FEI S P, CHEN Z, LI L, et al. Bayesian model averaging to improve the yield prediction in wheat breeding trials [J]. *Agricultural and Forest Meteorology*, 2023, 328: 109237.
- [6] FEI S P, XU D M, CHEN Z, et al. MLR-based feature splitting regression for estimating plant traits using high-dimensional hyperspectral reflectance data [J]. *Field Crops Research*, 2023, 293: 108855.
- [7] FEI S, XIAO Y, MA Y, et al. Interpretable machine learning for hyperspectral wheat yield prediction [J]. *Field Crops Research*, 2024, 306: 109201.
- [8] COTROZZI L, PERON R, TUINSTRA M R, et al. Spectral phenotyping of physiological and anatomical leaf traits related with maize water status [J]. *Plant Physiology*, 2020, 184(3): 1363.
- [9] NAVARRO A, NICASTRO N, COSTA C, et al. Sorting biotic and abiotic stresses on wild rocket by leaf-image hyperspectral data mining with an artificial intelligence model [J]. *Plant Methods*, 2022, 18(1): 45.
- [10] MAIMAITIJANG M, SAGAN V, SIDIKE P, et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning [J]. *Remote Sensing of Environment*, 2020, 237: 111599.
- [11] SIDIKE P, SAGAN V, QUMSIYEH M, et al. Adaptive trigonometric transformation function with image contrast and color enhancement: Application to unmanned aerial system imagery [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(3): 404.
- [12] SAGAN V, MAIMAITIJANG M, SIDIKE P, et al. Uav/satellite multiscale data fusion for crop monitoring and early stress detection [J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, XLII-2/W13: 715.
- [13] SHAH S H, ANGEL Y, HOUBORG R, et al. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat [J]. *Remote Sensing*, 2019, 11(8): 920.
- [14] SINGH H, ROY A, SETIA R, et al. Estimation of chlorophyll, macronutrients and water content in maize from hyperspectral data using machine learning and explainable artificial intelligence techniques [J]. *Remote Sensing Letters*, 2022, 13(10): 969.
- [15] ATLIN G N, CAIRNS J E, DAS B. Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change [J]. *Global Food Security*, 2017, 12: 31.
- [16] FU P, MEACHAM-HENSOLD K, GUAN K Y, et al. Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms [J]. *Frontiers in Plant Science*, 2019, 10: 730.
- [17] ZHOU J, LI B, XU X, et al. Lightweight 3D CNN for potato

[1] 魏益民, 张波, 关二旗, 等. 中国冬小麦品质改良研究进展[J]. 中国农业科学, 2013, 46(20): 4189.

WEI Y M, ZHANG B, GUAN E Q, et al. Advances in study

- yield estimation using UAV imagery [J]. *Precision Agriculture*, 2024, 25(2): 512.
- [18] WAN L, ZHOU W J, HE Y, *et al.* Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets [J]. *Remote Sensing of Environment*, 2022, 269: 112826.
- [19] SHU M Y, FEI S P, ZHANG B Y, *et al.* Application of UAV multisensor data and ensemble approach for high-throughput estimation of maize phenotyping traits [J]. *Plant Phenomics*, 2022, 2022: 9802585.
- [20] PIPIA L, AMIN E, BELDA S, *et al.* Green LAI and LCC Retrieval for pasture quality assessment using UAV-based hyperspectral data and gaussian processes [J]. *Remote Sensing*, 2021, 13(12): 2337.
- [21] CHEN R Z, ZHANG C, XU B, *et al.* A dynamic ensemble approach with feature selection for crop yield prediction [J]. *Computers and Electronics in Agriculture*, 2024, 217: 108567.
- [22] JI C L, ZHANG C, HUA L, *et al.* A multi-scale evolutionary deep learning model based on CEEMDAN, improved whale optimization algorithm, regularized extreme learning machine and LSTM for AQI prediction [J]. *Environmental Research*, 2022, 215: 114228.
- [23] SOLTANI A, HEYDARI M, AGHAEI F, *et al.* Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms [J]. *Cities*, 2022, 131: 103941.
- [24] WU P C, HE Y, LI Y F, *et al.* Online prediction of cutting temperature using self-adaptive local learning and dynamic CNN [J]. *IEEE Transactions on Industrial Informatics*, 2022, 18(12): 8629.
- [25] ZHANG F Y, SU X C, TAN A L, *et al.* Prediction of research octane number loss and sulfur content in gasoline refining using machine learning [J]. *Energy*, 2022, 261: 124823.
- [26] SUBHA S, BAGHAVATHI PRIYA S. Prediction model for a good learning environment using an ensemble approach [J]. *Computer Systems Science and Engineering*, 2023, 44(3): 2081.
- [27] CHEN R J, TSAY Y S, ZHANG T. A multi-objective optimization strategy for building carbon emission from the whole life cycle perspective [J]. *Energy*, 2023, 262: 125373.
- [28] FU P, MEACHAM-HENSOLD K, GUAN K Y, *et al.* Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms [J]. *Frontiers in Plant Science*, 2019, 10: 730.
- [29] FENG L W, ZHANG Z, MA Y C, *et al.* Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning [J]. *Remote Sensing*, 2020, 12(12): 2028.
- [30] CHEN R Q, ZHANG C J, XU B, *et al.* Predicting individual apple tree yield using UAV multi-source remote sensing data and ensemble learning [J]. *Computers and Electronics in Agriculture*, 2022, 201: 107275.
- [31] SHU M Y, DONG Q Z, FEI S P, *et al.* Improved estimation of canopy water status in maize using UAV-based digital and hyperspectral images [J]. *Computers and Electronics in Agriculture*, 2022, 197: 106982.
- [32] GUO T, LI H, YANG X, *et al.* Linking hyperspectral reflectance to grain filling dynamics in wheat: Implications for high-throughput phenotyping [J]. *Field Crops Research*, 2023, 291: 108765.
- [33] MAIMAITIJANG M, SAGAN V, SIDIKE P, *et al.* Vegetation spectral phenotyping for crop breeding: Combining UAV-based remote sensing and machine learning to assess genetic variation in wheat [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 171: 188.
- [34] SHU M Y, DONG Q Z, FEI S P, *et al.* Improved estimation of canopy water status in maize using UAV-based digital and hyperspectral images [J]. *Computers and Electronics in Agriculture*, 2022, 197: 106982.
- [35] FEI S P, XU D M, CHEN Z, *et al.* MLR-based feature splitting regression for estimating plant traits using high-dimensional hyperspectral reflectance data [J]. *Field Crops Research*, 2023, 293: 108855.
- [36] TIAN L, WANG Z Y, XUE B W, *et al.* A disease-specific spectral index tracks *Magnaporthe oryzae* infection in paddy rice from ground to space [J]. *Remote Sensing of Environment*, 2023, 285: 113384.
- [37] HERNANDEZ J, LOBOS G, MATUS I, *et al.* Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum aestivum* L.) grown under three water regimes [J]. *Remote Sensing*, 2015, 7(2): 2109.
- [38] ANSARIFAR J, WANG L Z, ARCHONTOULIS S V. An interaction regression model for crop yield prediction [J]. *Scientific Reports*, 2021, 11: 17754.
- [39] RAMAN R, KANTARI H, GOKHALE A A, *et al.* Agriculture yield estimation using machine learning algorithms [C]//2024 International Conference on Automation and Computation (AUTOCOM). Dehradun, India. IEEE, 2024: 187.
- [40] 郭燕, 李强, 张伟, 等. 基于高光谱成像的小麦灌浆期籽粒干物质积累监测研究 [J]. *农业工程学报*, 2023, 39(5): 150.
- GUO Y, LI Q, ZHANG W, *et al.* Monitoring dry matter accumulation in wheat grains during filling stage using hyperspectral imaging [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(5): 150.
- [41] FEI S, HASSAN M A, YANG X, *et al.* Comparative analysis of vegetation indices for monitoring phenological stages in wheat [J]. *Field Crops Research*, 2021, 263: 108075.
- [42] WANG X G, LI Y, ZHANG R X, *et al.* Multi-temporal hyperspectral imaging with deep ensemble learning for winter wheat yield forecasting [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 207: 112.
- [43] LIANG M, CHANG T P, AN B X, *et al.* A stacking ensemble learning framework for genomic prediction [J]. *Frontiers in Genetics*, 2021, 12: 600040.
- [44] ZHANG R, ZHAO L, WANG Q, *et al.* Hyperspectral and LiDAR data fusion for crop monitoring: A feature-level ensemble approach [J]. *Remote Sensing of Environment*, 2024, 304: 113402.
- [45] LI H, ZHAO J, WANG L, *et al.* 3D plant modeling enhanced maize yield prediction by fusing UAV LiDAR and hyperspectral data [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 209: 154.
- [46] ZHANG Y, LIU X, CAO J, *et al.* Integrating multi-temporal Sentinel-2 and hyperspectral data for rice yield prediction using stacked generalization [J]. *Field Crops Research*, 2023, 302: 109034.
- [47] MAIMAITIJANG M, SAGAN V, SIDIKE P, *et al.* Transfer learning for hyperspectral crop yield prediction across species and environments [J]. *Nature Food*, 2024, 5(3): 210.
- [48] NEJAD S M M, ABBASI-MOGHADAM D, SHARIFI A, *et al.* Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 16: 254.