

# 基于无人机图像和 SHAP 特征筛选的小麦田间产量预测方法研究

朱志畅<sup>1</sup>, 葛焱<sup>1,2</sup>, 臧晶荣<sup>3</sup>, 李庆<sup>3</sup>, 金时超<sup>4</sup>, 徐焕良<sup>1</sup>, 翟肇裕<sup>1</sup>

(1. 南京农业大学人工智能学院, 江苏南京 210031; 2. 南京农业大学工学院, 江苏南京 210031;

3. 南京农业大学农学院, 江苏南京 210095; 4. 南京农业大学前沿交叉研究院, 江苏南京 210095)

**摘要:** 为了探寻适宜的小麦产量预测模型并提高其精度, 从冬小麦灌浆期的无人机多光谱和 RGB 图像中提取了 14 种光谱参数和 28 种形态参数作为特征变量, 利用线性回归、随机森林、神经网络等 10 种机器学习方法构建小麦田间产量预测模型, 并比较了模型间预测能力的差异; 同时, 引入机器学习事后可解释性方法 SHAP 对输入的特征变量进行重要性分析和筛选, 了解其提高模型预测能力的效果。结果表明: (1) 10 种机器学习模型中, 误差逆传播神经网络 BPNN 的产量预测表现最好 ( $r^2=0.826$ ,  $RMSE=0.094 \text{ t} \cdot \text{hm}^{-2}$ ); (2) 根据 SHAP 确定的特征变量重要性排序, 花青素反射指数 ARI 和三维冠层体积 Volume 对于预测结果的影响最大, 占全部特征重要性总和的 45.48%; (3) 经过 SHAP 特征筛选后, 确定了在 BPNN 产量预测模型上表现最优的 9 个特征变量, 其预测结果  $r^2$  为 0.865,  $RMSE$  为  $0.075 \text{ t} \cdot \text{hm}^{-2}$ , 比使用全特征的 BPNN 和事前 Pearson 相关性分析方法在预测精度上均有提升。因此, 在优选产量预测模型基础上, 可采用 SHAP 机制对特征变量的重要性进行筛选和分析, 以此进一步提高田间小麦产量预测精度。

**关键词:** 小麦; 无人机图像; 机器学习; SHAP 加性解释方法; 产量预测

中图分类号: TP391; S512.1+1

文献标识码: A

文章编号: 1009-1041(2025)02-0264-11

## Research on Wheat Yield Prediction Based on UAV Imagery and SHAP Feature Selection

ZHU Zhichang<sup>1</sup>, GE Yan<sup>1,2</sup>, ZANG Jingrong<sup>3</sup>, LI Qing<sup>3</sup>, JIN Shichao<sup>4</sup>, XU Huanliang<sup>1</sup>, ZHAI Zhaoyu<sup>1</sup>

(1. College of Artificial Intelligence, Nanjing Agricultural University, Nanjing, Jiangsu 210031, China; 2. College of Engineering, Nanjing Agricultural University, Nanjing, Jiangsu 210031, China; 3. College of Agriculture, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China; 4. Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China)

**Abstract:** Accurate and effective yield prediction is essential for wheat breeding, cultivation and field management. In this study, the multispectral and RGB images of winter wheat during the grain filling stage were collected from UAV, and 14 spectral traits and 28 morphological traits were extracted as feature variables. Ten machine learning methods, including linear regression, random forest and neural network, were used to construct wheat yield prediction models, and the differences between the models were compared to select the best one. Additionally, machine learning interpretability method SHAP was introduced to analyze the importance of the feature variables, in order to improve the prediction performance of the model. The results showed that among the 10 machine learning methods used, the BPNN model had the best prediction performance ( $r^2=0.826$ ,  $RMSE=0.094 \text{ t} \cdot \text{hm}^{-2}$ ). According to the feature importance ranking determined by SHAP, Anthocyanin Reflectance Index (ARI) and Three-Dimensional Canopy Volume (Volume) had the greatest impact on the prediction results, accounting for 45.48% of the total feature importance. After feature selection using SHAP,

收稿日期: 2024-01-15 修回日期: 2024-02-29

基金项目: 江苏省自然科学基金项目(BK20231004); 中央高校基本科研业务费专项(KYCXJC2023007)

第一作者 E-mail: 30217331@njau.edu.cn(朱志畅); geyan@njau.edu.cn(葛焱)

通讯作者 E-mail: zhaoyu.zhai@njau.edu.cn(翟肇裕)

the BPNN model with the best performance was determined based on nine feature variables ( $r^2 = 0.865$ ,  $RMSE = 0.075 \text{ t} \cdot \text{hm}^{-2}$ ). This improved the prediction accuracy compared to the BPNN model using all features and the pre-analysis method Pearson correlation analysis. Therefore, based on the optimal yield prediction model, SHAP mechanism can be used to select and analyze the importance of feature variables, so as to further improve the accuracy of wheat yield prediction.

**Keywords:** Wheat; UAV imagery; Machine learning; Shapley additive explanations; Yield prediction

及时、准确、无损地预测作物产量一直是农业领域备受关注的重点课题,随着无人机遥感技术的快速发展,越来越多的农业科研人员利用无人机遥感数据开展作物产量预测,在小麦<sup>[1]</sup>、棉花<sup>[2]</sup>、玉米<sup>[3]</sup>等作物上逐渐被推广应用。其中,最广泛使用的遥感数据为 RGB 图像,其能够很好地反映作物的颜色、纹理和结构特征<sup>[4]</sup>,且对设备要求相对较低,获取手段也较为方便。另一种则是多光谱图像,其优势在于能够反映作物丰富的光谱信息,这些信息对于田间作物的长势和生理状况都有着很好的监测效果<sup>[5]</sup>。

当前,研究人员已经从 RGB 图像中找到了预测作物生物量和产量的有效技术流程。例如,从无人机 RGB 图像中提取小麦的颜色指数(CIs)和小麦冠层高度信息作为估算地上部生物量的特征变量,能够显著提高估算精度<sup>[6]</sup>;采用无人机获取蚕豆小区的 RGB 图像,结合单一机器学习和集成学习方法,对蚕豆的地上部生物量和产量进行预测,其中基于集成学习的产量预测模型表现更佳( $r^2 = 0.854$ ,  $RMSE = 0.568 \text{ t} \cdot \text{hm}^{-2}$ )<sup>[7]</sup>。与此同时,通过多光谱图像预测产量的相关研究也日臻成熟。如,Guo 等<sup>[8]</sup>利用无人机多光谱图像提取了玉米田块的植被指数和纹理指数,分别构建了估测叶绿素含量、叶面积指数和产量的 5 种机器学习模型和 5 种深度学习模型,找到了估测上述三类指标的最佳模型;Li 等<sup>[9]</sup>从无人机多光谱图像中提取 16 个产量敏感植被指数,结合不同机器学习方法构建冬小麦产量预测模型,预测结果的误差在  $-0.1 \sim 0.1 \text{ t} \cdot \text{hm}^{-2}$  之间。还有部分研究将无人机 RGB 与多光谱图像中提取的作物表型信息融合,试图从提高特征丰富度的角度来增强产量预测效果。如,Herrero 等<sup>[10]</sup>从无人机 RGB 图像生成三维点云中提取了最大高度、冠层体积等形态参数,并从无人机多光谱图像中提取了 NDVI、SAVI、MSAVI 等植被指数,将两类特征同时输入到机器学习模型中预测大豆产量,模型精确度(ACC)达到 91.36%;Fei 等<sup>[11]</sup>将无人

机 RGB 和多光谱数据融合,利用 5 种机器学习方法构建小麦产量预测模型,得出数据融合提高了预测精度的结论( $RMSE = 0.916 \text{ t} \cdot \text{hm}^{-2}$ )。

然而,不管是使用单一信息还是多源信息进行作物产量预测,机器学习模型的性能都不可避免地会达到一定瓶颈,一个主要原因就是机器学习模型在可解释性上的匮乏<sup>[12]</sup>。在很多情况下,使用者一般直接将全部特征变量输入到模型之中,这就导致了一些与预测任务相关性低、重要性弱的无关变量对模型的正常学习产生干扰,降低了预测效果。为此,部分学者探索利用相关性分析等方法确定输入特征的重要性并进行筛选,以提高预测能力。如,Liu 等<sup>[13]</sup>利用 Pearson 相关性分析确定了与产量相关性较强的 3 个植被指数,结合 3 种机器学习方法构建小麦产量预测模型,表现最佳模型的  $r^2$ 、RMSE 和 MAE 分别为 0.85、793.96  $\text{kg} \cdot \text{hm}^{-2}$ 、656.31  $\text{kg} \cdot \text{hm}^{-2}$ ;Wang 等<sup>[14]</sup>利用顺序向前选择(SFS)对哨兵 2A 卫星(Sentinel-2A)获取莫索湾棉花垦区光谱图像的不同波段和植被指数进行特征重要性筛选,构建棉花产量预测模型,绘制了总体精度较高( $r^2 = 0.62$ ,  $RE = 11.06\%$ )的大尺度棉花产量预测分布图。SHAP(Shapley additive explanations)是一种解释机器学习模型输出的方法,它通过 SHAP 值对模型的预测结果进行解释<sup>[15]</sup>。周亚男等<sup>[16]</sup>利用 SHAP 对基于 Stacking 集成学习的山地丘陵区土地覆被分类模型进行解释,为山地丘陵区土地覆被分类任务的特征变量选择提供了新的思路。葛建坤等<sup>[17]</sup>利用 SHAP 结合八种机器学习算法对灌区渠道调度流量进行预测,筛选出 4 个重要特征结合随机森林算法进行预测,预测精度有效提高。SHAP 作为一种机器学习事后可解释性方法,在利用机器学习解决实际问题的应用研究中已被使用<sup>[18-19]</sup>。但目前,较少有学者在农作物特别是小麦的田间产量预测上使用该方法。

本研究以冬小麦为对象,采集了灌浆期的无

人机 RGB 和多光谱图像,分别提取了以植被指数为代表的光谱参数和以冠层高度、三维体积等为代表的形态参数作为特征变量,比较了不同机器学习算法构建的小麦产量预测模型的表现,并引入 SHAP 机制对输入模型的特征变量进行重要性分析和筛选,以期能够进一步提高小麦产量预测的精度。

## 1 材料与方 法

### 1.1 研究区域和试验设计

研究区位于江苏省南京市溧水区南京农业大

学白马教学科研基地(119°10'37"E,31°37'09"N),试验在 2021 年 10 月至 2022 年 5 月开展。溧水区气候属于北亚热带季风气候,年平均降水量 1 204.3 mm。

田间试验以 56 个小麦品种为供试材料,采用随机区组设计,设置 0、120 和 240 kg·hm<sup>-2</sup> 三个施氮水平(分别用 N0、N120 和 N240 表示),3 次重复,小区面积为 1.2 m<sup>2</sup>(1.2 m×1 m),行间距 0.2 m。田间布局如图 1 所示。小麦整个生育期内适时灌水,及时进行病虫草害管理,保证植株的正常生长。

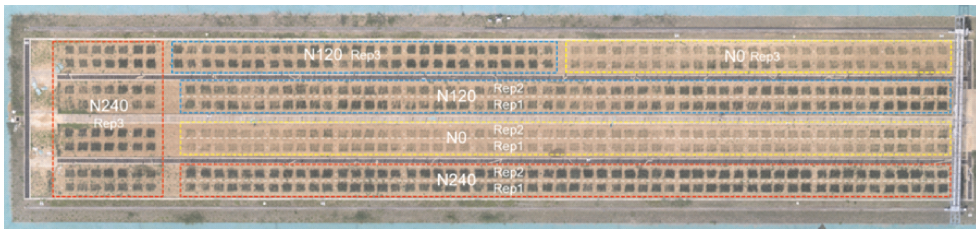


图 1 试验设计

Fig. 1 Experimental design

### 1.2 数据采集

#### 1.2.1 无人机图像采集

多光谱图像采用八旋翼无人机 Ecodrone UAS-8 搭载五通道多光谱相机进行采集,包含蓝光波段(450±16)nm、绿光波段(560±16)nm、红光波段(650±16)nm、红边波段(730±16)nm 和近红外波段(840±16)nm,分辨率为 1 280×960 像素;飞行任务开始前,在地面放置光谱校正板,对图像进行辐射校正。RGB 图像使用大疆精灵 4 无人机采集,为方便后续的图像拼接、三维重构等数据处理工作,选择 Pix4Dcapture 软件进行航线规划和航拍参数设置;飞行方式采用交叉网格,范围为 120 m×35 m,高度为 20 m,相机倾斜角度为 80°,正面和侧面重叠度均为 80%。

灌浆期是小麦籽粒物质积累和产量形成的关键生育时期<sup>[20]</sup>,也是无人机遥感预测小麦产量的最佳时期<sup>[21]</sup>。因此,本研究选用冬小麦灌浆期的 RGB 和多光谱图像进行产量预测分析。数据采集工作选定于 2022 年 4 月 24 日(播种后第 160 天)正午进行,当天晴朗无云,光照条件较好,共采集原始多光谱图像 2 835 张和 RGB 图像 216 张。

#### 1.2.2 产量数据获取

在小麦完全成熟后,人工完成收割,并在阳光下晾晒、脱粒后进行测产。

### 1.3 数据处理

#### 1.3.1 图像预处理

RGB 图像采集完毕后,依次完成配准、拼接、几何校正等预处理操作。利用 Pix4Dmapper 软件的三维点云重构模块,生成试验区域的点云数据。再根据小区实际坐标,对试验区域点云进行分割,获得全部 504 个小区的独立点云数据,并继续进行点云去噪、滤波和归一化处理等标准化操作<sup>[22]</sup>。

多光谱图像预处理方法与 RGB 图像大致相同,但需要增加辐射校正和波段融合两个步骤。根据采集图像中的漫反射地面光谱校正板,结合各波段的光谱反射系数,使用 Pix4Dmapper 软件中的辐射校正模块完成校正。再使用 ENVI 5.1 遥感图像处理软件进行波段融合,以便后续光谱参数的提取。

#### 1.3.2 表型参数提取

光谱参数选择 ARI、CIgreen、CVI 等 14 个植被指数进行研究(表 1)。所选植被指数主要分为两类:一类是与作物冠层叶片色素含量相关的指数,如 CIgreen、CVI、NCPI 与叶绿素含量相关,ARI 与花青素含量相关;另一类与作物形态结构和功能密切相关,如 NDVI、NIRv、PSRI 等。

表 1 植被指数  
Table 1 Vegetation indices

植被指数 Vegetation index	计算公式 Equation	参考文献 Reference
花青素反射指数 ARI	$\frac{1}{G} - \frac{1}{N_{ir}}$	[23]
绿色叶绿素指数 CIgreen	$\frac{N_{ir}}{G} - 1$	[24]
叶绿素植被指数 CVI	$\frac{N_{ir} \times R}{G^2}$	[25]
差异植被指数 DVI	$N_{ir} - R$	[26]
增强植被指数 EVI	$\frac{2.5 \times (N_{ir} - R)}{N_{ir} + 6 \times R - 7.5 \times B + 1}$	[27]
绿色归一化差异植被指数 GNDVI	$\frac{N_{ir} - G}{N_{ir} + G}$	[28]
改进土壤调节植被指数 MSAVI	$\frac{N_{ir} - R}{N_{ir} + R + L} (1 + L), L = 0.1$	[29]
归一化叶绿素色素植被指数 NCPI	$\frac{R - B}{R + B}$	[30]
归一化差异植被指数 NDVI	$\frac{N_{ir} - R}{N_{ir} + R}$	[31]
归一化绿红差异植被指数 NGRDI	$\frac{G - R}{G + R}$	[31]
近红外反射率植被指数 NIRv	$N_{ir} \times \frac{N_{ir} - R}{N_{ir} + R}$	[32]
植被衰老反射指数 PSRI	$\frac{R - G}{N_{ir}}$	[33]
比值植被指数 RVI	$\frac{N_{ir}}{R}$	[34]
结构不敏感色素指数 SIPI	$\frac{N_{ir} - B}{N_{ir} - R}$	[35]

计算公式中  $B, G, R, N_{ir}$  分别表示在蓝光波段(450 nm)、绿光波段(560 nm)、红光波段(650 nm)、近红外波段(840 nm)的小麦冠层光谱反射率。

$B, G, R$  and  $N_{ir}$  in the equations represent the spectral reflectance of the wheat canopy in blue (450 nm), green (560 nm), red (650 nm) and near-infrared (840 nm), respectively.

考虑到冠层形态结构展现维度的差异,从各小区点云中提取包含一维、二维、三维共计 28 种形态参数(表 2)进行研究。其中,最大高度(maximum height)、平均高度(mean height)、高度百分位数(height quantiles)等一维参数与作物株高相关;冠层覆盖度(canopy cover)、植物面积指数(plant area index)、投影叶面积(projected leaf area)等二维参数与作物叶面积相关;三维冠层体积(3D canopy volume)、三维体积指数(3D volume index)、三维轮廓指数(3D profile index)等三维参数与作物体积相关,这些参数在过去的研究中被证明对于作物的形态结构、生理状态有着很好的反映能力<sup>[36]</sup>。

#### 1.4 模型构建

选用 10 种常见的机器学习方法来构建产量预测模型。其中,线性回归算法(多元线性回归 MLR 和逐步多元线性回归 SMR)是形式较为简单且易于建模的基础算法;基于最小二乘思想的回归算法(偏最小二乘 PLSR 和岭回归 RDR)能

够很好地处理多变量共线性问题;基于决策树的算法(决策树 DT、极限梯度提升回归 XGBoost 和随机森林 RF)则是利用树形结构的思想,其中每个内部结点表示一种属性的判断,在有效完成任务的同时还具备较小的计算开销;K 近邻算法是一种基于距离理念的非线性学习模型,时间复杂度较低,但空间复杂度较高;支持向量回归 SVR 则是一种基于计算学习理论的算法,在预测任务较复杂时能保持较高的有效性,但算力要求也相对更大;误差逆传播神经网络 BPNN 是一种基本的人工前馈神经网络模型,具有处理非线性关系、高度稳定、容错能力等优点。

将提取的 14 个光谱参数和 28 个形态参数进行归一化处理,作为模型输入的特征变量,对 10 种机器学习模型预测小麦产量的效果进行对比。从数据集中随机选取 70% 为训练集,30% 为测试集,并进行随机交叉验证,以对模型的预测结果进行合理评估。

表 2 形态参数  
Table 2 Morphological traits

类别 Class	形态参数 Morphological trait	简称 Abbreviation
一维参数 1D trait	最大高度 Maximum height	Hmax
	平均高度 Mean height	Hmean
	高度百分位数 Height quantiles	H80-H99
二维参数 2D trait	冠层覆盖度 Canopy cover	CC
	植物面积指数 Plant area index	PAI
	投影叶面积 Absolute projected leaf area	PLA
三维参数 3D trait	三维冠层体积 3D canopy volume	Volume
	三维体积指数 3D volume index	3DVI
	三维轮廓指数 3D profile index	3DPI

### 1.5 基于 SHAP 的特征变量筛选方法

本研究在对传统的机器学习模型择优的基础上,引入 SHAP 机制对模型进行解释,将输入模型的特征变量进行重要性排序,筛选出最优特征变量组合,确定最佳产量预测模型。

SHAP 是一种解释机器学习模型输出的方法,属于根据预期表示一致性和局部准确性的加性特征归因方法<sup>[17]</sup>。SHAP 为模型的每一个特征变量分配一个特定的预测重要性值即 SHAP 值,作为特征变量重要性的统一度量。其将最优分配资源和局部解释联系起来,提供了基于 SHAP 值聚合的全局解释方法,可以在多特征变量交叉作用的环境下,解析每个特征对于预测结果的贡献度<sup>[16]</sup>。SHAP 值的计算可表示为以下等式:

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (1)$$

其中, $g$  为解释模型, $z' \in \{0,1\}^M$  表示相应特征变量是否存在, $M$  是输入的整体特征变量的数量, $\phi_i \in R$  代表第  $i$  个变量的边际贡献, $\phi_0$  是所有训练集的预测均值,作为解释模型的常数。进而,可以将 SHAP 值的具体求解表示为以下等式:

$$Y_n = y_b + f(x_n, 1) + f(x_n, 2) + \dots + f(x_n, P) \quad (2)$$

其中, $Y_n$  为所求 SHAP 值; $y_b$  为所求特征变量在所有样本上的均值; $f(x_n, 1)$  是第  $n$  个样本中第 1 个特征变量的贡献值,依此类推; $P$  为特征

变量的数量。

### 1.6 模型评价

为了评估模型性能,选择决定系数 (correlation of determination,  $r^2$ )、均方根误差 (root mean square error, RMSE) 作为模型预测精度主要评价指标。 $r^2$  是评价模型拟合优度常用的统计指标,表示因变量的变异性能够由模型解释的比例。RMSE 则可以衡量模型预测的误差大小,并且与实际预测指标单位相同。当  $r^2$  越接近 1, RMSE 越小时,认定为模型的预测精度更高。同时,还选择了平均绝对误差 (mean absolute error, MAE) 和相对误差 (relative error, RE) 作为评价模型表现的辅助指标,以对模型预测效果进行更加全面地分析和判断。MAE 和 RE 的绝对值越小,模型的预测精度越高。

$$r^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i| \quad (5)$$

$$RE = \frac{1}{m} \sum_{i=1}^m \frac{|\hat{y}_i - y_i|}{y_i} \quad (6)$$

其中, $\hat{y}_i$  表示预测产量, $y_i$  表示实测产量, $\bar{y}$  是  $y_i$  的均值, $m$  是样本的数量。

## 2 结果与分析

### 2.1 不同机器学习模型产量预测效果比较

不同机器学习模型在产量预测中的精度表现有所不同(表 3),测试集  $r^2$  和 RMSE 的变化范围分别为 0.628~0.826 和 0.086~0.130 t·hm<sup>-2</sup>。总体来看,集成学习模型(如 XGBoost、RF)的表现优于单一模型(如 DT);逐步多元回归 SMR( $r^2=0.806$ )的表现优于多元线性回归 MLR( $r^2=0.628$ );BPNN 的  $r^2$  最大,达到 0.826;RF 的 RMSE 最小,仅 0.086 t·hm<sup>-2</sup>;MLR 的  $r^2$  和 RMSE 都是所有模型中最差的。由于 BPNN 在  $r^2$  上表现明显优于其他模型,在 RMSE 及其他误差指标上接近最低值,选择 BPNN 进行后续研究与分析。

### 2.2 基于 SHAP 的特征变量筛选

#### 2.2.1 特征变量重要性分析

为进一步掌握各特征变量对产量预测的重要性,使用 SHAP 方法对输入 BPNN 的特征变量进

表 3 不同机器学习模型产量预测对比

Table 3 Yield prediction evaluation of different machine learning models

模型 Model	训练集 Training set				测试集 Test set			
	$r^2$	RMSE/ ( $t \cdot hm^{-2}$ )	MAE/ ( $t \cdot hm^{-2}$ )	RE	$r^2$	RMSE/ ( $t \cdot hm^{-2}$ )	MAE/ ( $t \cdot hm^{-2}$ )	RE
多元线性回归 MLR	0.711	0.110	0.085	0.417	0.628	0.130	0.093	0.494
K 近邻回归 KNR	0.748	0.096	0.068	0.255	0.707	0.108	0.075	0.349
决策树 DT	0.726	0.094	0.071	0.265	0.717	0.109	0.077	0.362
岭回归 RDR	0.738	0.096	0.071	0.329	0.722	0.108	0.075	0.349
偏最小二乘回归 PLSR	0.828	0.084	0.073	0.406	0.767	0.102	0.078	0.435
极限梯度提升 XGBoost	0.855	0.076	0.061	0.126	0.804	0.090	0.069	0.223
逐步多元回归 SMR	0.892	0.068	0.058	0.189	0.806	0.090	0.068	0.274
随机森林 RF	0.966	0.039	0.054	0.156	0.813	0.086	0.063	0.190
支持向量回归 SVR	0.863	0.073	0.066	0.143	0.817	0.095	0.069	0.208
误差逆传播网络 BPNN	0.862	0.073	0.061	0.19	0.826	0.094	0.066	0.265

行可解释性分析。从图 2 可以看出,虽然众多特征变量参与了模型的运算,但真正对产量预测发挥作用的变量并不多。图 2a 的光谱参数中,ARI 起到了主要作用,其对作物叶片中的花青素含量较为敏感<sup>[37]</sup>,而花青素可以通过吸收光能减少叶绿素对光的吸收,起到潜在的动态调节作物光合作用的效果<sup>[38]</sup>。CVI 对于作物的冠层叶绿素含量有着很好的指示作用<sup>[39]</sup>。另外,RVI 和 GNDVI 的平均绝对 SHAP 值为 0,说明它们在模型中几乎没有作用。RVI 在植被高密度覆盖的情况下较为敏感,在植被覆盖度低于 50% 时分辨能力显著下降<sup>[40]</sup>。GNDVI 同样对于植被覆盖度较为敏感。本试验中,部分中低施氮水平小区的植被覆盖度不高,这可能是造成 RVI 和 GNDVI 重要性低的原因之一。

图 2b 形态参数重要性排序中,Volume 和 Hmean 的重要性凸显,PAI、PLA、H99、H97 等也发挥了一定的作用,这些参数主要反映了作物的冠层高度、体积等结构信息,在作物形态结构的表征上优于一般的植被指数<sup>[41]</sup>。其中,Volume 是三维冠层体积,通过将 3D 点体素化计算出冠层体积,目前已被证明与产量密切相关<sup>[20]</sup>。H85、H82 等高度百分位数的重要性普遍靠后,并未很好地反映出小麦的冠层结构,与产量相关性不高。

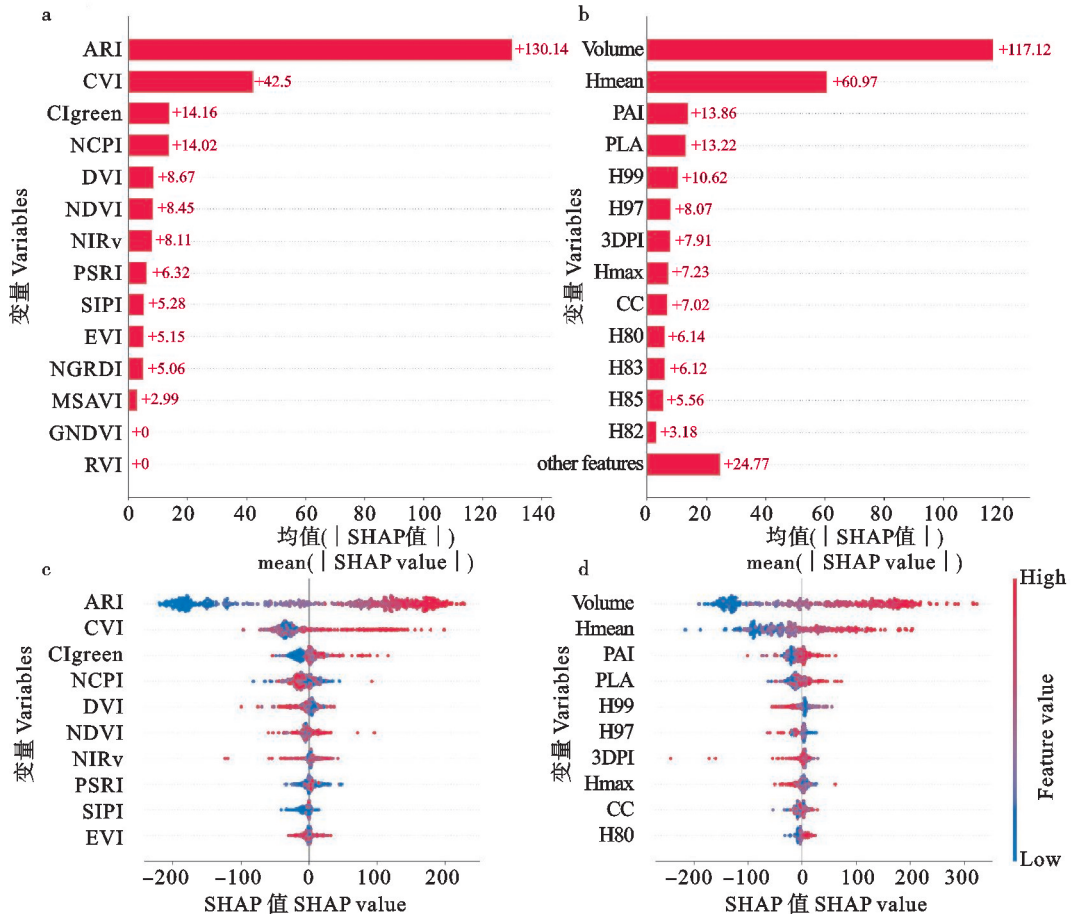
蜂群图(图 2c 和图 2d)可以很好地反映每个特征变量的 SHAP 值分布情况,还可揭示特征变量的取值与 SHAP 值之间的关系,这是 SHAP 可解释性的直观展现<sup>[42]</sup>。图 2c 中,ARI 取值越大,对模型输出的正影响越大;取值越小,对模型输出的负影响越大。然而,CVI 对模型输出的负影响不随其取值变小而下降,CIgreen 也展现出同样

的趋势,说明 CVI 和 CIgreen 只有在取值较大时与产量呈正相关,取值较小时相关性不高。NCPI 则与产量呈负相关,即取值越大,反而对模型的输出产生了一定的负影响,取值越小却产生了正影响。图 2d 中,Volume 和 Hmean 对于模型输出的影响最大,且都为正相关。部分与高度有关的特征变量(如 H99、H97、Hmax)呈负相关,但不显著。

根据 SHAP 的解释结果,对特征变量的重要性进行排序(表 4)。其中,重要性排名前 3 的特征变量( $X_1 \sim X_3$ )累计贡献度超过 50%,前 4( $X_1 \sim X_4$ )的累计贡献度超过 60%,前 7( $X_1 \sim X_7$ )的累计贡献度超过 70%,前 12( $X_1 \sim X_{12}$ )的累计贡献度超过 80%,前 19( $X_1 \sim X_{19}$ )的累计贡献度超过 90%。前 16 位特征变量累计贡献度为 86.83%,接近 90%,且  $X_{16}$  的 SHAP 值大于 7,贡献度大于 1.2%,重要性明显高于其后变量,可以认定  $X_1 \sim X_{16}$  在模型中发挥了主要作用。

### 2.2.2 特征变量组合筛选

为进一步优化特征筛选,对重要性排名靠前的 16 个特征变量作为模型输入,逐步剔除变量组合中重要性较低的变量(从  $X_{16}$  开始)。从不同变量组合下 BPNN 模型的产量预测效果(表 5)看,以 16 个特征变量作为输入组合的产量预测模型测试集  $r^2$  为 0.845, RMSE 为  $0.082 t \cdot hm^{-2}$ ;随着特征变量的逐步剔除,测试集  $r^2$  总体呈现先升后降趋势, RMSE 和 MAE 大体呈先降后升趋势;当变量组合为  $X_1 \sim X_9$  时,  $r^2$  达到最大值(0.865), RMSE 和 MAE 分别达到最小值(0.075 和  $0.057 t \cdot hm^{-2}$ ),模型预测效果达到最优,即  $X_1 \sim X_9$  的 9 特征组合为当前最佳组合。



a 和 c: 光谱参数; b 和 d: 形态参数。  
a and c: spectral traits; b and d: morphological traits.

图 2 SHAP 可解释性分析结果  
Fig. 2 SHAP interpretability analysis result

表 4 SHAP 特征变量重要性排序

Table 4 Feature importance ranking of SHAP

符号 Symbol	特征变量 Feature variable	SHAP 绝对值 SHAP absolute value	贡献度 Contribution/%	累计贡献度 Cumulative contribution/%
X <sub>1</sub>	ARI	130.14	23.94	23.94
X <sub>2</sub>	Volume	117.12	21.54	45.48
X <sub>3</sub>	Hmean	60.97	11.21	56.69
X <sub>4</sub>	CVI	42.5	7.82	64.51
X <sub>5</sub>	CIgreen	14.16	2.60	67.11
X <sub>6</sub>	NCPI	14.02	2.58	69.69
X <sub>7</sub>	PAI	13.86	2.55	72.24
X <sub>8</sub>	PLA	13.22	2.43	74.67
X <sub>9</sub>	H99	10.62	1.95	76.63
X <sub>10</sub>	DVI	8.67	1.59	78.22
X <sub>11</sub>	NDVI	8.45	1.55	79.78
X <sub>12</sub>	NIRv	8.11	1.49	81.27
X <sub>13</sub>	H97	8.07	1.48	82.75
X <sub>14</sub>	3DPI	7.91	1.45	84.21
X <sub>15</sub>	Hmax	7.23	1.33	85.54
X <sub>16</sub>	CC	7.02	1.29	86.83
X <sub>17</sub>	PSRI	6.32	1.16	87.99
X <sub>18</sub>	SIPI	6.32	1.16	89.15
X <sub>19</sub>	H80	6.14	1.13	90.28
X <sub>20</sub>	H83	6.12	1.13	91.41

贡献度为该特征变量的 SHAP 值在所有变量 SHAP 值总和中所占比例。  
Contribution refers to the proportion of the SHAP absolute value of a variable in the sum of SHAP absolute values of all variables.

表 5 不同特征变量组合预测效果分析

Table 5 Analysis of model accuracy with different combinations of features

特征变量组合 Feature group	训练集 Training set			测试集 Test set		
	$r^2$	RMSE/ ( $t \cdot \text{hm}^{-2}$ )	MAE/ ( $t \cdot \text{hm}^{-2}$ )	$r^2$	RMSE/ ( $t \cdot \text{hm}^{-2}$ )	MAE/ ( $t \cdot \text{hm}^{-2}$ )
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}+X_{13}+X_{14}+X_{15}+X_{16}$	0.861	0.081	0.057	0.845	0.082	0.058
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}+X_{13}+X_{14}+X_{15}$	0.866	0.080	0.059	0.847	0.086	0.060
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}+X_{13}+X_{14}$	0.869	0.068	0.060	0.845	0.080	0.062
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}+X_{13}$	0.872	0.073	0.053	0.847	0.084	0.062
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}+X_{12}$	0.873	0.065	0.054	0.849	0.081	0.063
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}$	0.875	0.059	0.054	0.854	0.078	0.062
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}$	0.886	0.061	0.057	0.855	0.078	0.060
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9$	0.898	0.063	0.052	0.865	0.075	0.057
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8$	0.879	0.064	0.053	0.849	0.085	0.062
$X_1+X_2+X_3+X_4+X_5+X_6+X_7$	0.862	0.077	0.055	0.845	0.082	0.058
$X_1+X_2+X_3+X_4+X_5+X_6$	0.868	0.067	0.059	0.829	0.086	0.064
$X_1+X_2+X_3+X_4+X_5$	0.883	0.070	0.055	0.827	0.087	0.058
$X_1+X_2+X_3+X_4$	0.862	0.073	0.058	0.826	0.094	0.064
$X_1+X_2+X_3$	0.802	0.090	0.061	0.754	0.105	0.081
$X_1+X_2$	0.764	0.113	0.065	0.686	0.117	0.086
$X_1$	0.691	0.100	0.073	0.657	0.118	0.086

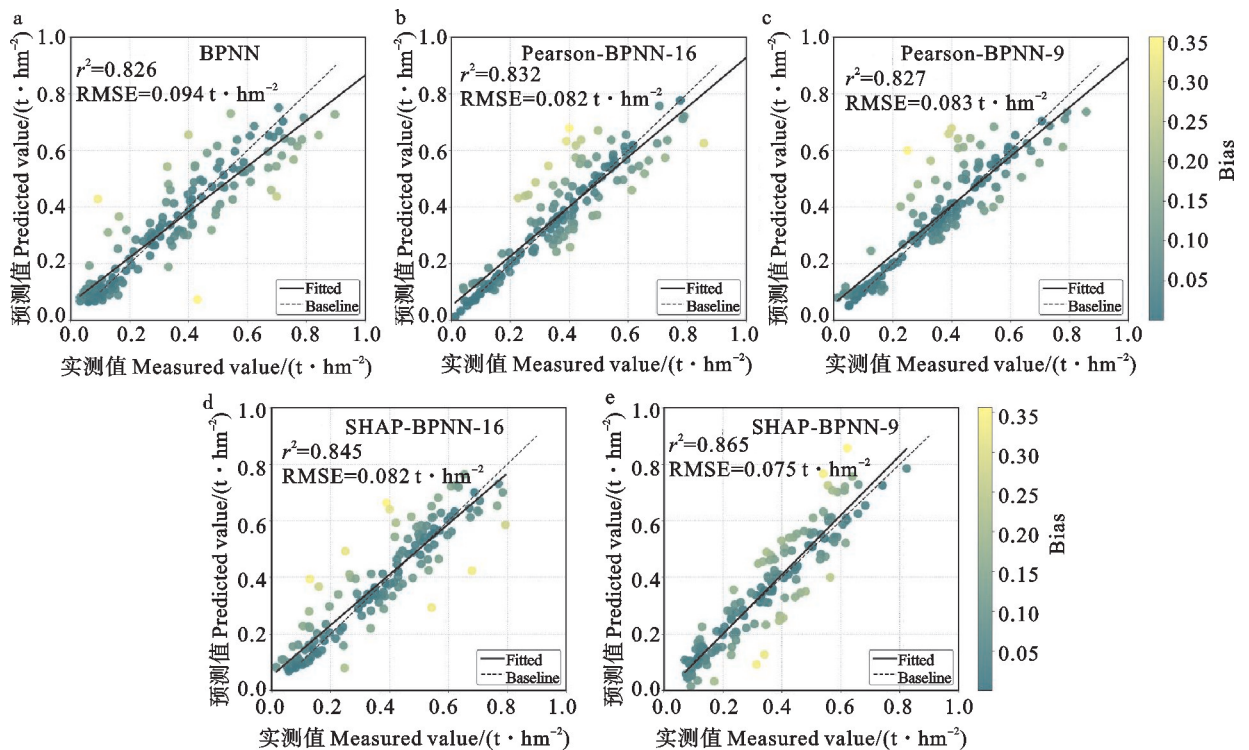


图 3 不同模型预测值与真实值相关性散点图

Fig. 3 Correlation between measured yield and predicted yield of different models

### 2.3 基于 SHAP 特征筛选的预测效果比较

为进一步验证基于 SHAP 的事后特征筛选与传统的事前特征筛选方法的差异,选用 Pearson 相关性分析法进行对比,结果(表 6)表明,不

管是 Pearson 还是 SHAP,经过特征筛选后的模型预测效果普遍优于全特征的预测效果;同样,不管是 16 特征组合还是 9 特征组合,SHAP-BPNN 的预测效果均优于 Pearson-BPNN。其中,Pearson-

BPNN-16 表现略优于 Pearson-BPNN-9; 然而, SHAP-BPNN-9 比 SHAP-BPNN-16 展现出更好的预测精度,  $r^2$  最高达 0.865, 比全特征组合高出约 0.04, 比 16 特征组合高出 0.02, RMSE 比全特征组合和 16 特征组合分别低 0.019 和 0.007  $t \cdot hm^{-2}$ 。

图 3 为 Pearson 和 SHAP 的不同特征组合预测值与真实值的相关性散点图。可以进一步看出, SHAP-BPNN 比 BPNN 的预测结果更加接近实际产量, 且出现误差较大的预测(即离群点)的情况较少。另外, BPNN 在实际产量较大时, 预测产量准确性普遍偏低, 引入 SHAP 机制后则有效缓解了这一偏向, 使得预测结果更加精确。

表 6 不同特征筛选方法的预测效果分析  
Table 6 Prediction analysis of different feature selection methods

模型 Model	$r^2$	RMSE/ ( $t \cdot hm^{-2}$ )	MAE/ ( $t \cdot hm^{-2}$ )	RE
BPNN	0.826	0.094	0.066	0.265
Pearson-BPNN-16	0.832	0.082	0.065	0.250
Pearson-BPNN-9	0.827	0.083	0.053	0.223
SHAP-BPNN-16	0.845	0.082	0.058	0.270
SHAP-BPNN-9	0.865	0.075	0.057	0.198

### 3 讨论

#### 3.1 关于不同机器学习模型在产量预测上的应用

不同机器学习模型的底层思想和算法原理不同, 所以适用场景和优缺点也不同。在产量预测任务中, 多元线性回归模型( $r^2 = 0.628$ )表现较差, 可能由于特征变量数较多且各具备丰富的含义, 导致产量预测成为一个复杂的非线性问题, 线性模型难以从众多特征变量中进行有效的学习, 这也是集成学习模型表现优于单一模型的原因之一。然而, 逐步多元回归 SMR( $r^2 = 0.806$ )却展现出了较好的预测效果, 可能与其在模型运算过程中使用了 AIC 值和  $F$  检验自动剔除了部分特征变量有关。同时, BPNN 的表现之所以优于其他模型, 主要原因在于神经网络结构具有天然的非线性特征学习能力和拟合能力, 并且还使用了误差逆传播算法, 克服了隐藏层连接权重学习问题, 能够更好地通过调整权重和偏置来优化模型, 使得模型达到最佳状态。

#### 3.2 关于不同特征变量筛选方法在产量预测上的应用

在实际应用中, 不同特征变量对于机器学习

模型所需完成任务的贡献不同, 因而选择合适的特征变量作为输入是提高模型性能的有效步骤。目前, 大多数研究聚焦在使用模型之前, 对输入的特征变量进行相关性分析、筛选和降维, 常用方法包括 Pearson 相关性分析、PCA 主成分分析、方差阈值法等。这些方法都是在模型完成实际任务前对特征变量进行筛选, 具体筛选出的特征变量是否能很好地在模型中发挥作用, 只能通过模型最终的结果来体现。被剔除的变量是否真的对完成任务无关, 也无法得到肯定答案。而 SHAP 作为一种事后可解释性方法, 工作机制是通过模型的运算结果反推输入到模型中的各个不同特征变量对于模型输出的贡献度。研究表明, 通过 SHAP 筛选的特征变量能提高模型的预测效果, 且优于通过 Pearson 相关性分析筛选的特征变量预测结果, 进一步印证了事后可解释性方法的优越性。

#### 3.3 关于利用 SHAP 机制筛选重要表型参数的问题

经过 SHAP 筛选的特征变量组合, 不管是 16 个还是 9 个特征组合, 都既包含了反映光谱信息的植被指数, 又包含了反映结构信息的形态参数, 说明多源数据融合对于小麦产量的预测具有积极作用, 与 Han 等<sup>[43]</sup>的研究结果一致。具体来看, 植被指数中 ARI(23.94%)、CVI(7.82%)、CI-green(2.60%)、NCPI(2.58%) 等反映作物内部色素含量的指数重要性排名较高, 这与 Li 等<sup>[41]</sup>研究结果一致。形态参数中 20 个高度百分位数的重要性排名总体靠后, 只有 H99 进入了 9 特征组合之中, 说明高度百分位数在小区级产量预测任务中并未发挥出足够的作用, 这与 Jin 等<sup>[36]</sup>在玉米小区产量预测上有所不同。究其原因, 一方面高度百分位数是对作物不同高度层结构特征的精准反映, Jin 等研究玉米植株相较于小麦植株在体型上更加高大(尤其是进入生长发育后期), 高度百分位数能够反映的特征相对来说更加丰富, 对于产量的响应也就更加明显; 另一方面, 本试验小区面积设置较小(1.2  $m^2$ ), 植株之间存在相互交叠情况, 重构的三维数据更加难以精准地反映作物的内部结构状况。

### 4 结论

相比于线性回归、偏最小二乘、随机森林等机器学习方法, BPNN 具有较好的小麦产量预测效

果( $r^2=0.826$ ,  $RMSE=0.094 \text{ t} \cdot \text{hm}^{-2}$ )。在此基础上,引入 SHAP 机制对 BPNN 的预测结果进行有效解释,并量化每个特征变量在模型预测中发挥的作用。选取筛选出的最佳特征变量组合(9 特征)作为输入变量,BPNN 模型  $r^2$  达 0.865、 $RMSE$  为  $0.075 \text{ t} \cdot \text{hm}^{-2}$ ,比使用全特征组合在  $r^2$  上提升了 0.039, $RMSE$  上下降了  $0.019 \text{ t} \cdot \text{hm}^{-2}$ ,比 Pearson 相关性分析法不管在 16 特征还是 9 特征组合中,均表现出更优的产量预测能力。

#### 参考文献:

- [1] 甘甜,李雷,李红叶,等.基于多源遥感数据和机器学习算法的冬小麦产量预测研究[J].麦类作物学报,2022,42(11):1419.  
GAN T, LI L, LI H Y, et al. Winter wheat yield prediction based on multi-source remote sensing data and machine learning algorithms [J]. *Journal of Triticeae Crops*, 2022, 42(11):1419.
- [2] 胡灿,王兴旺,王旭峰,等.基于 Android 手机的田间棉花产量预测系统设计[J].农业机械学报,2023,54(S2):252.  
HU C, WANG X W, WANG X F, et al. Field cotton yield prediction system based on android mobile phone [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(S2):252.
- [3] RIGDEN A J, MUELLER N D, HOLBROOK N M, et al. Combined influence of soil moisture and atmospheric evaporative demand is important for accurately predicting US maize yields [J]. *Nature Food*, 2020, 1:127.
- [4] SU J, ZHU X, LI S, et al. AI meets UAVs: A survey on AI empowered UAV perception systems for precision agriculture [J]. *Neurocomputing*, 2023, 518(C):242.
- [5] 张羽,杨涛,马吉锋,等.数学形态学辅助下基于光谱指数的作物冠层组分类[J].农业工程学报,2022,38(7):163.  
ZHANG Y, YANG T, MA J F, et al. Classification of crop canopy components based on spectral index assisted by mathematical morphology [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(7):163.
- [6] WANG D, LI R, ZHU B, et al. Estimation of wheat plant height and biomass by combining UAV imagery and elevation data [J]. *Agriculture*, 2022, 13(1):9.
- [7] JI Y, LIU R, XIAO Y, et al. Faba bean above-ground biomass and bean yield estimation based on consumer-grade unmanned aerial vehicle RGB images and ensemble learning [J]. *Precision Agriculture*, 2023, 24(4):1439.
- [8] GUO Y, XIAO Y, HAO F, et al. Comparison of different machine learning algorithms for predicting maize grain yield using UAV-based hyperspectral images [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2023, 124:103528.
- [9] LI Z, CHEN Z, CHENG Q, et al. Deep learning models outperform generalized machine learning models in predicting winter wheat yield based on multispectral data from drones [J]. *Drones*, 2023, 7(8):505.
- [10] HERRERO-HUERTA M, RODRIGUEZ-GONZALVEZ P, RAINEY K M. Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean [J]. *Plant Methods*, 2020, 16(1):78.
- [11] FEI S, HASSAN M A, XIAO Y, et al. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat [J]. *Precision Agriculture*, 2023, 24(1):187.
- [12] CASTELVECCHI D. Can we open the black box of AI? [J]. *Nature*, 2016, 538:20.
- [13] LIU J, ZHU Y, TAO X, et al. Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-grade unmanned aerial vehicles multispectral imagery [J]. *Frontiers in Plant Science*, 2022, 13:1032170.
- [14] 王汇涵,张泽,康孝岩,等.基于 Sentinel-2A 的棉花种植面积提取及产量预测[J].农业工程学报,2022,38(9):205.  
WANG H H, ZHANG Z, KANG X Y, et al. Cotton planting area extraction and yield prediction based on Sentinel-2A [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(9):205.
- [15] LUNDBERG S M, LEE S-I. A unified approach to interpreting model predictions [C]. In *Advances in Neural Information Processing Systems*, 2017:4765.
- [16] 周亚男,陈绘,刘洪斌.基于多源数据和 Stacking-SHAP 方法的山地丘陵区土地覆被分类[J].农业工程学报,2022,38(23):213.  
ZHOU Y N, CHEN H, LIU H B. Land cover classification in hilly and mountainous areas using multi-source data and Stacking-SHAP technique [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(23):213.
- [17] 葛建坤,雷国相,陈皓锐,等.基于 SHAP 重要性排序和机器学习算法的灌区渠道调度流量预测[J].农业工程学报,2023,39(13):113.  
GE J K, LEI G X, CHEN H R, et al. Irrigation district channel dispatch flow prediction based on SHAP importance ranking and machine learning algorithm [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(13):113.
- [18] ZHANG J, MA X, ZHANG J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model [J]. *Journal of Environmental Management*, 2023, 332:117357.
- [19] CHEHREH CHELGANI S, NASIRI H, TOHRY A, et al. Modeling industrial hydrocyclone operational variables by SHAP-CatBoost - A "conscious lab" approach [J]. *Powder Technology*, 2023, 420:118416.
- [20] 施凌天,丁国辉,夏云鹏,等.结合激光雷达和三维性状分析的田间小麦产量分级研究[J].南京农业大学学报,2023,46

- (6):1012.  
SHI L T, DING G H, XIA Y P, *et al.* Combining backpack LiDAR and 3D trait analysis to enable in-field yield classification in wheat [J]. *Journal of Nanjing Agricultural University*, 2023, 46(6):1012.
- [21] 杨楠, 周萌, 陈欢, 等. 基于无人机 RGB 影像的小麦叶面积指数与产量估算研究[J]. *麦类作物学报*, 2023, 43(7):920.  
YANG N, ZHOU M, CHEN H, *et al.* Estimation of wheat leaf area index and yield based on UAV RGB images [J]. *Journal of Triticeae Crops*, 2023, 43(7):920.
- [22] SU Y, WU F, AO Z, *et al.* Evaluating maize phenotype dynamics under drought stress using terrestrial lidar [J]. *Plant Methods*, 2019, 15(1):11.
- [23] GITELSON A A, MERZLYAK M N, CHIVKUNOVA O B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves [J]. *Photochemistry and Photobiology*, 2001, 74(1):38.
- [24] GITELSON A A, VERMA S B, VIÑA A, *et al.* Novel technique for remote estimation of CO<sub>2</sub> flux in maize [J]. *Geophysical Research Letters*, 2003, 30(9):1486.
- [25] BLACKMER T M, SCHEPERS J S, VARVEL G E. Light reflectance compared with other nitrogen stress measurements in corn leaves [J]. *Agronomy Journal*, 1994, 86(6):934.
- [26] RICHARDSON A, WIEGAND C. Distinguishing vegetation from soil background information [J]. *Photogrammetric Engineering & Remote Sensing*, 1978, 43:1.
- [27] HUETE A, DIDAN K, MIURA T, *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices [J]. *Remote Sensing of Environment*, 2002, 83(1-2):195.
- [28] GITELSON A A, KAUFMAN Y J, MERZLYAK M N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS [J]. *Remote Sensing of Environment*, 1996, 58(3):289.
- [29] QI J, CHEHBOUNI A, HUETE A R, *et al.* A modified soil adjusted vegetation index [J]. *Remote Sensing of Environment*, 1994, 48(2):119.
- [30] PEÑUELAS J, GAMON J A, FREDEEN A L, *et al.* Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves [J]. *Remote Sensing of Environment*, 1994, 48(2):135.
- [31] TUCKER C J. Red and photographic infrared linear combinations for monitoring vegetation [J]. *Remote Sensing of Environment*, 1979, 8(2):127.
- [32] BADGLEY G, FIELD C B, BERRY J A. Canopy near-infrared reflectance and terrestrial photosynthesis [J]. *Science Advances*, 2017, 3(3):e1602244.
- [33] MERZLYAK M N, GITELSON A A, CHIVKUNOVA O B, *et al.* Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening [J]. *Physiologia Plantarum*, 1999, 106(1):135.
- [34] JORDAN C F. Derivation of leaf-area index from quality of light on the forest floor [J]. *Ecology*, 1969, 50(4):663.
- [35] PENUÉLAS J, BARET F, FILELLA I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance [J]. *Photosynthetica*, 1995, 31(2):221.
- [36] JIN S, SU Y, SONG S, *et al.* No-destructive estimation of field maize biomass using terrestrial lidar: an evaluation from plot level to individual leaf level [J]. *Plant Methods*, 2020, 16:69.
- [37] 梁守真, 隋学艳, 王猛, 等. 花青素对植物反射特性的影响及遥感估算: 叶片尺度[J]. *光谱学与光谱分析*, 2024, 44(1):275.  
LIANG S Z, SUI X Y, WANG M, *et al.* The influence of anthocyanin on plant optical properties and remote sensing estimation at the scale of leaf [J]. *Spectroscopy and Spectral Analysis*, 2024, 44(1):275.
- [38] 王伟东, 常庆瑞, 王玉娜. 冬小麦叶片花青素相对含量高光谱监测[J]. *麦类作物学报*, 2020, 40(6):754.  
WANG W D, CHANG Q R, WANG Y N. Hyperspectral monitoring of anthocyanins relative content in winter wheat leaves [J]. *Journal of Triticeae Crops*, 2020, 40(6):754.
- [39] CLEVERS J, KOOISTRA L, VAN DEN BRANDE M. Using sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop [J]. *Remote Sensing*, 2017, 9(5):405.
- [40] 尹芬, 丁美青. 运用 ENVI 实现 SPOT-5 卫星影像的 RVI 提取[J]. *测绘与空间地理信息*, 2016, 39(1):47.  
YIN F, DING M Q. RVI extraction of SPOT-5 satellite images using ENVI [J]. *Geomatics & Spatial Information Technology*, 2016, 39(1):47.
- [41] LI Q, JIN S, ZANG J, *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.
- [42] LUNDBERG S. SHAP documentation: beeswarm plot [EB/OL]. (2024-12-04) [2024-01-14]. [https://shap.readthedocs.io/en/latest/example\\_notebooks/api\\_examples/plots/beeswarm.html](https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html).
- [43] HAN J, ZHANG Z, CAO J, *et al.* Prediction of winter wheat yield based on multi-source data and machine learning in China [J]. *Remote Sensing*, 2020, 12(2):236.