

# 基于高光谱反射率估测分蘖洋葱叶片净光合速率的研究

韩青妍<sup>1</sup>, 崔洪博<sup>1</sup>, 刘燕妮<sup>1</sup>, 王秀峰<sup>1</sup>, 张泽锦<sup>2</sup>, 王剑锋<sup>1\*</sup>

(1. 吉林省蔬菜花卉科学研究院/吉林省设施蔬菜重点实验室, 长春 130119; 2. 四川省农业科学园艺研究所, 成都 610066)

**摘要:**分蘖洋葱的叶片为筒状叶, 常规的光合作用测量仪无法使用其叶室夹取叶片, 准确获取净光合速率。为解决这一问题, 利用高光谱反射率结合机器学习算法来定量分析分蘖洋葱叶片的净光合速率。采用主成分分析(PCA)与随机森林(RF)模型分别和支持向量机(SVM)模型相结合, 从高光谱数据中有效地提取了反映净光合速率的光谱特征, 建立了高光谱反射率反演净光合速率的模型。主成分分析有效降低了数据维数, 捕获了前9个主成分92.9%的方差, 为机器学习模型提供了坚实的基础。在测试的模型中, 随机森林模型表现出色, 决定系数( $R^2$ )为0.94, 均方根误差(RMSE)为1.5154, 预测精度和稳定性都很高。尽管支持向量机模型的 $R^2$ 为0.81, RMSE为1.63, 但仍被证明能够处理数据中复杂的非线性关系, 展示了其在预测建模任务中的鲁棒性和泛化能力。比较分析发现, 随机森林模型更适合构建净光合速率的高光谱反射率反演模型。它能够通过集成策略减少预测错误, 提供更稳定、更准确的估测。因此, 本研究所采用的方法是解决分蘖洋葱筒状叶片净光合速率估测的有效手段。

**关键词:**光谱分析; 净光合速率; 支持向量机模型; 随机森林模型

中图分类号: S633.2

文献标识码: A

文章编号: 2096-5877(2024)06-0083-06

## Research on Estimating the Net Photosynthetic Rate of Tillering Onion Leaves Based on Hyperspectral Reflectance

HAN Qingyan<sup>1</sup>, CUI Hongbo<sup>1</sup>, LIU Yanni<sup>1</sup>, WANG Xiufeng<sup>1</sup>, ZHANG Zejin<sup>2</sup>, WANG Jianfeng<sup>1\*</sup>

(1. Jilin Provincial Institute of Vegetable and Flower Science/Jilin Provincial Key Laboratory of Protected Vegetable, Changchun 130119; 2. Institute of Horticulture, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China)

**Abstracts:** The leaves of tillering onions are tubular, and conventional photosynthesis measurement instruments are unable to accurately gauge the net photosynthetic rate using their leaf chamber clamps. To address this issue, this paper utilizes hyperspectral reflectance combined with machine learning algorithms to quantitatively analyze the net photosynthetic rate in the leaves of tillering onions. Principal Component Analysis (PCA) was employed in conjunction with Random Forest (RF) and Support Vector Machine (SVM) models to effectively extract spectral features indicative of the net photosynthetic rate from hyperspectral data, establishing a model for the inversion of net photosynthetic rate based on hyperspectral reflectance. Principal Component Analysis effectively reduced the dimensionality of the data, capturing 92.9% of the variance within the first nine principal components, thereby providing a solid foundation for the machine learning models. In the tested models, the Random Forest model performed excellently, with a coefficient of determination ( $R^2$ ) of 0.94 and a root mean square error (RMSE) of 1.5154, demonstrating high prediction accuracy and stability. Although the Support Vector Machine model had a lower  $R^2$  of 0.81 and an RMSE of 1.63, it was still proven capable of handling complex nonlinear relationships in the data, showing robustness and generalization ability in predictive modeling tasks. Comparative analysis indicates that the Random Forest model is more suited for constructing models for the inversion of net photosynthetic rate using hyperspectral reflectance. This model reduces predictive errors through an ensemble strategy, providing more stable and accurate estimates. There-

收稿日期: 2024-08-08

基金项目: 吉林省科技发展计划项目(20240601079RC); 国家现代农业产业技术体系项目(CARS-24)

作者简介: 韩青妍(1998-), 女, 研究实习员, 硕士, 主要从事设施蔬菜植物生理检测研究。

通信作者: 王剑锋, 男, 博士, 副研究员, E-mail: hortwjf@163.com

fore, the methodology employed in this study represents an effective approach for estimating the net photosynthetic rate in the tubular leaves of tillering onions.

**Key words:** Spectral analysis; Net photosynthetic rate; Support Vector Machine; Random Forest

分蘖洋葱是我国东北地区广泛种植与消费的蔬菜之一,因其独特的风味、营养价值以及多种用途,在农业产业中占据重要地位<sup>[1-3]</sup>。分蘖洋葱不仅具有抗炎和抗氧化特性,同时还作为食品加工原料<sup>[4]</sup>。因此,在带动地方经济,促进农民增收方面具有重要意义<sup>[5]</sup>。

在农业生产中有效获取作物的净光合速率值对生产管理至关重要<sup>[6]</sup>。然而,目前的光合作用测量仪仅能检测片状叶片和针状叶片的净光合速率(Pn),无法检测具有筒状叶特点的分蘖洋葱叶片。在不破坏分蘖洋葱叶片的前提下,光合作用测量仪的叶室无法夹取其筒状叶片,无法形成完全密封,导致气体交换测量不准确,从而无法准确检测净光合速率。此外,分蘖洋葱叶片表面气孔分布的异质性,使得用标准方法测量净光合速率更加复杂化。这些问题凸显了使用光合作用测量仪测量非标准植物叶片存在的局限性。

利用高光谱反射率反演作物叶片的净光合速率是一种快速、无损有效的技术手段<sup>[7]</sup>。近年来,随着遥感技术的不断进步,高光谱反射率反演光合参数已成为研究热点<sup>[8]</sup>。这种技术为作物生理生化指标的定量检测带来了更多的可能性,越来越多的研究应用高光谱技术手段来检测作物光合参数<sup>[9-10]</sup>。在小麦旗叶的高光谱反射率波段中,利用一阶导数光谱变换后提取敏感波段,并采用BP神经网络法(BPNN)、偏最小二乘法(PLSR)和二次多项式逐步回归(QPSR)三种方法构建净光合速率Pn的反演模型,比较了三种模型的预测精度,结果显示,这些建模方法和敏感波段提取对于Pn的反演预测是可行的<sup>[11]</sup>。在水稻叶片光谱和光合入射有效辐射的研究中,通过使用叶绿素相关的植被指数与光合入射有效辐射的乘积来反演叶片的净光合速率,结果表明,这种方法可以显著提高模型的精度, $R^2$ 均大于0.75,满足反演要求<sup>[12]</sup>。在估算冬小麦叶片净光合速率的研究中,采用小波变换对高光谱数据进行分解,分析不同尺度下光谱指数与净光合速率的相关性,提取最佳分解层次的敏感波段构建光谱指数,并建立BP神经网络模型,建模结果显示RMSE为1.99, $R^2$ 为0.86<sup>[13]</sup>。董德誉等<sup>[14]</sup>采用梯度增强回归和岭回归方法,建立了正常施氮处理下小麦开花期和全生

育期四个光合参数的估算模型,结果表明,梯度增强回归能够较好地预测施氮下开花期的净光合速率,决定系数( $R^2$ )为0.82。王娣等<sup>[15]</sup>在分析光合有效辐射、植被光谱指数与净光合速率Pn关系的基础上,建立了Pn的反演模型,结果表明,比值植被指数 $RVI_{700}$ 、红边归一化植被指数 $NDVI_{705}$ 和红边叶绿素指数 $CI_{rededge}$ 这三种植被指数与叶绿素含量相关性较高, $R^2$ 最高达到0.88,当结合光合有效辐射和植被光谱指数的乘积来分析Pn时,相关性显著提高,尤其是 $CI_{rededge}$ 和光合有效辐射的乘积与Pn的反演效果最佳。上述研究采用高光谱反射率估测作物片状叶片净光合速率取得了良好的效果。

然而,目前采用高光谱反射率建立模型,估测分蘖洋葱筒状叶片净光合速率的研究较少。为此,本研究将主成分分析法和随机森林分别与支持向量机结合,建立分蘖洋葱叶片净光合速率高光谱反射率反演模型,并对比研究基于不同方法构建的高光谱反射率反演模型的精度,获得最优反演模型,从而为分蘖洋葱叶片Pn快速获取和无损监测提供依据。

## 1 材料与方 法

### 1.1 试验材料

本研究选用“金龙珠”分蘖洋葱作为样本,培育试验区位于吉林省蔬菜花卉科学研究所,分蘖洋葱的种植时间为2024年4月上旬至2024年7月中旬。采用日本园试配方营养液,浓度为1.0、2.0、3.0、4.0 mS/cm四个处理,从而培育出不同净光合速率的分蘖洋葱叶片,2024年5月至6月每隔7 d滴灌一次营养液,每次用量为300 mL/株。每个处理种植500株分蘖洋葱,样本采集时间为2024年5月至6月,在不同处理的试验区,随机采集叶片,共采集100片。

### 1.2 高光谱数据采集

分蘖洋葱叶片高光谱反射率数据的测定在实验室内进行。采用光谱仪(AvaSpec-ULS2048XL-EVO)测定。将光纤接收端与余弦校正器相连,并固定在操作台上以保持水平状态。使用调节型标准钨灯(Ls-181106)作为光源,并在光纤连接点增设红外隔板以降低红外波段的光强影响,防止过

强的光照对叶片造成灼伤。光源与余弦校正器之间保持固定距离,并设置成45°夹角。叶片平放在该夹角的正下方。在测量期间,关闭暗室的操作窗口以避免外界光源的干扰,确保能够测得更准确的高光谱反射率。测量波长范围为400~1100 nm,步长间隔为1 nm。

### 1.3 叶片内外表皮细胞观察

选取植株中部新鲜叶片,将叶片平整摆放在操作台上,使用无菌刀片分别撕取分蘖洋葱叶片内外表皮(大小5 mm×5 mm),放在超景深显微系统(VHX-7000)下观察,放大倍数为200倍。分蘖洋葱叶片表皮细胞上的气孔呈不规则圆形,大小和形状相对一致,而叶片内部则没有气孔分布,这些特征与普通的叶肉细胞在形态上有明显的区别。根据叶片气孔分布,可确定刨开后叶片在光合作用测量系统叶室中的夹取方向。

### 1.4 光合指标测定

为保证试验环境的稳定性,测量期间实验室内温度设为24℃,湿度为60%,模拟环境光强设为800 μmol/(m<sup>2</sup>·s)稳定预处理,采用光合作用测量仪(LI-6800, LICOR公司,美国)测量净光合速率。待测试完反射光谱后,用刀片将分蘖洋葱叶片划开,平整地放入叶室内,待数值稳定后记录分蘖洋葱叶片准确净光合速率数值。测试样本与高光谱反射率测试样本一致。

### 1.5 模型精度评价指标

为了评估建立的高光谱反射率反演模型的准确率,本研究选取了两种模型对高光谱反射率反演净光合速率模型。本文详细介绍了使用支持向量机和随机森林建立的模型在验证集上的性能评估方法和过程。通过输入高光谱反射率数据来反演净光合速率。

评估方法包括均方根误差(RMSE)和决定系数(R<sup>2</sup>)<sup>[16]</sup>。计算公式如下

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{act} - y_{pred})^2} \dots\dots\dots (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{act} - y_{pred})^2}{\sum_{i=1}^n (y_{act} - \bar{y}_{act})^2} \dots\dots\dots (2)$$

式中 y<sub>act</sub> 为实测值; y<sub>pred</sub> 为预测值; n 为样本量; RMSE 为预测值与实际值之间差异的平方和的均值的平方根,反映了预测误差的大小。决定系数(R<sup>2</sup>)反映了模型对观测数据的拟合程度,其值越接近1,表示拟合程度越好。

## 1.6 数据统计与分析

利用Excel 2013、Origin 2018等软件对原始数据进行处理。利用MATLAB R2021b软件分析高光谱数据。

## 2 结果与分析

### 2.1 叶片光合指标数据统计特征

试验共测量100组数据,叶片净光合速率数据见图1。分蘖洋葱叶片净光合速率范围为0.51~12.67 μmol/(m<sup>2</sup>·s),平均值为7.41 μmol/(m<sup>2</sup>·s),标准差为3.33 μmol/(m<sup>2</sup>·s),样本分布符合正态分布,可用作后续研究。

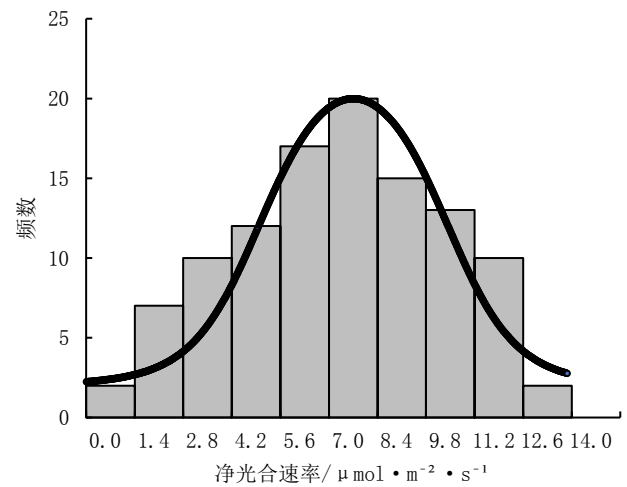


图1 分蘖洋葱叶片净光合速率分布

### 2.2 叶片光谱特征

分蘖洋葱叶片高光谱反射率曲线如图2所示。在获取分蘖洋葱叶片高光谱反射率数据基础上进行主成分分析时,首先需要对原始高光谱反射率数据进行平滑处理,本研究用Origin软件中

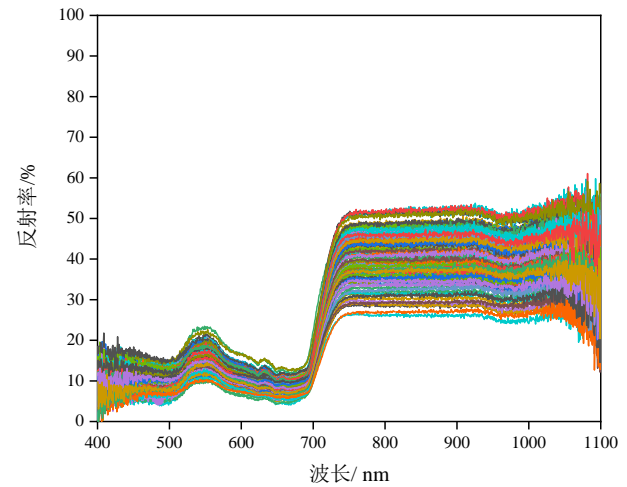


图2 分蘖洋葱叶片高光谱反射率曲线

的Savitzky-Golay方法对原始光谱反射率数据进行处理,窗口点数为10,在保证光谱特征的前提下,消除光谱中存在的误差,平滑曲线如图3所示。

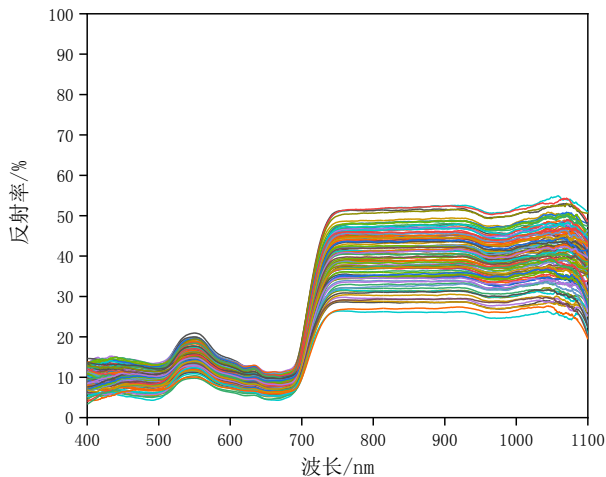


图3 平滑后的高光谱反射率曲线

从图3可以观察到,在400~700 nm的可见光波段,植物叶片的光谱特性是其光合作用效率和生理活动的关键表征。在蓝光区域(400~500 nm),反射率显著较低,表明此波段的光能被叶绿素a和b大量吸收。叶绿素对蓝光的高吸收率对于植物的光合作用及其生长发育至关重要,蓝光还影响植物的形态建成,如促进茎伸长和叶片展开。绿光区域(500~600 nm),反射率略有上升。这一现象表明绿光被叶绿素较少吸收,使得大部分绿光被反射,这也是分蘖洋葱叶片呈绿色的原因之一。尽管绿光的直接吸收不及蓝光或红光,但能更深入地穿透叶片,促进底层叶绿体的光合作用,对整体光能利用效率有所贡献。在红光区域(600~700 nm),反射率再次下降,显示出与蓝光区类似的低反射率模式。红光的这一特性表明其是光合作用中非常有效的光区,主要由于叶绿素对红光有强烈的吸收,有助于能量的高效转换和光合作用的进行。

综上所述,通过精确测量和分析不同光波段的高光谱反射率,可以深入理解分蘖洋葱的光合动态和生理反应。分蘖洋葱叶片在这些波长范围内有效地利用光能进行光合作用。在700~1100 nm的近红外区域,高光谱反射率显著增加,说明叶片色素对绿光发生一定的反射,对蓝光和红光吸收作用强烈。

本研究通过分析分蘖洋葱叶片的高光谱反射率,揭示了其在不同波长下的吸收和反射行为,对优化光合作用具有重要意义。

### 2.3 PCA降维与特征信息提取

在模型建立过程中,遵循统计学原则,选择合适数量的主成分至关重要。选择过少的主成分可能导致重要变量信息的丢失,而过多则可能引入高光谱反射率数据中的噪声,从而导致模型产生过拟合现象,影响预测准确性。为了确定最佳的主成分数,本研究采用了交叉验证的方法。在主成分累计可信度增量较小时,选择尽可能少的主成分数,以确保模型的简洁和效率。高光谱反射率进行PCA降维后,前10个主成分的累计可信度如图4所示。基于分蘖洋葱叶片高光谱反射率数据进行的主成分分析(PCA),得到了每个主成分对总数据变异解释度的累计可信度。其中,前9个主成分的累计可信度达92.9%,接近全面地解释了数据中的变异。

利用100个样本,以PCA降维得到前9个主成分作为输入变量,在MATLAB R2021b中分别建立基于随机森林模型和支持向量机的回归估算模型。

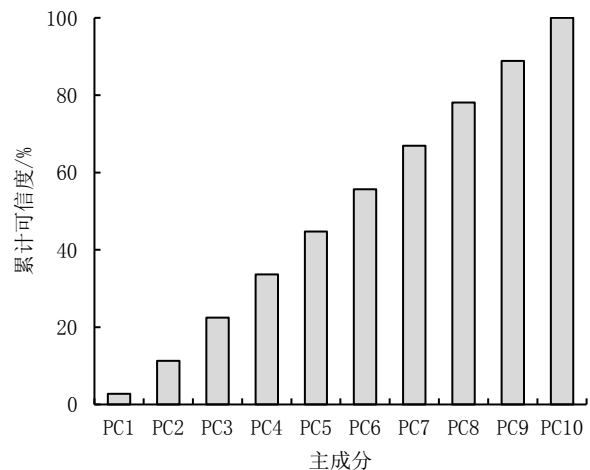


图4 前10个主成分的累计可信度

### 2.4 支持向量机模型的构建

支持向量机(SVM)是一种强大的监督学习方法,广泛应用于分类和回归问题中<sup>[6]</sup>。它的核心思想基于寻找一个最优的超平面,以此将不同类别的数据分隔开,达到分类的目的。在建立反演模型的过程中,采用了支持向量机(Support Vector Machine, SVM)算法。通过对训练集中的样本进行模型训练,得到一个葱叶净光合速率值与高光谱数据之间的映射关系。基于支持向量机(SVM)算法建立的分蘖洋葱叶片高光谱反射率反演净光合速率模型如图5所示。

从图5中可见,尽管数据与拟合方程存在一

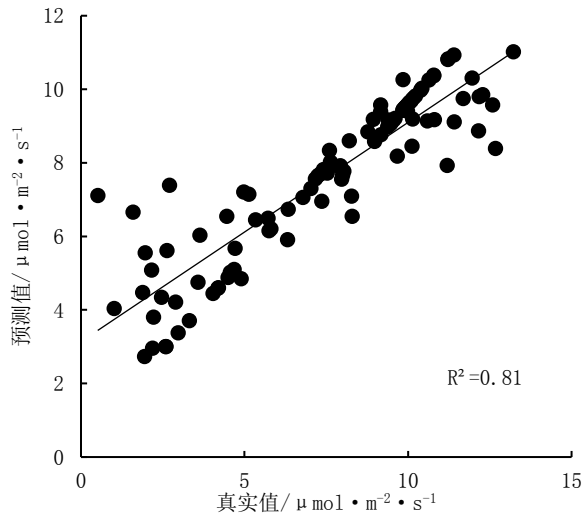


图5 支持向量机反演模型

定程度的偏差,支持向量机模型在分蘖洋葱叶片净光合速率的预测上具有较高的准确性。模型的决定系数( $R^2$ )为0.81,表示模型可以解释81%的数据变异性,而均方根误差(RMSE)为1.63,反映出模型预测值与实测值之间的平均误差程度。

### 2.5 随机森林模型的构建

随机森林模型是一种强大的集成学习方法<sup>[17]</sup>,通过构建多个决策树并结合其预测结果,有效提高了模型的准确性和稳定性,且降低了过拟合风险<sup>[18]</sup>。其核心原理包括自助采样生成数据子集、随机选择特征进行分裂,以及聚合各树的预测结果(分类取投票结果,回归取平均值)。这种策略引入了样本和特征的随机性,增强了模型的多样性和泛化能力,使随机森林广泛应用于分类、回归等复杂任务,展现出良好的性能和实用性<sup>[19]</sup>。构建随机森林反演模型如图6所示,该图展示了分蘖洋葱叶片光谱反演净光合速率的随机森林模型的预测结果与实际测量值之间的关系。横轴表示真实值,纵轴表示预测值。数据点分布显示了模型预测值与真实值之间的对应关系。从中可以观察到,数据点紧密地围绕对角线排列,表明预测值与真实值之间有很高的一致性。

模型的性能指标包括决定系数( $R^2$ )和均方根误差(RMSE),分别为0.94和1.5154。 $R^2$ 值为0.94说明模型能够解释94%的数据变异,显示出非常高的预测准确性。而RMSE值为1.5154,表示预测值与实际值之间的平均误差程度。为研究提供了一个清晰的视觉证据,证明所用的随机森林模型在分蘖洋葱叶片高光谱反射率反演方面表现出色。

以上统计指标说明,两种算法在处理具有复杂非线性关系的数据集方面均表现出良好的性

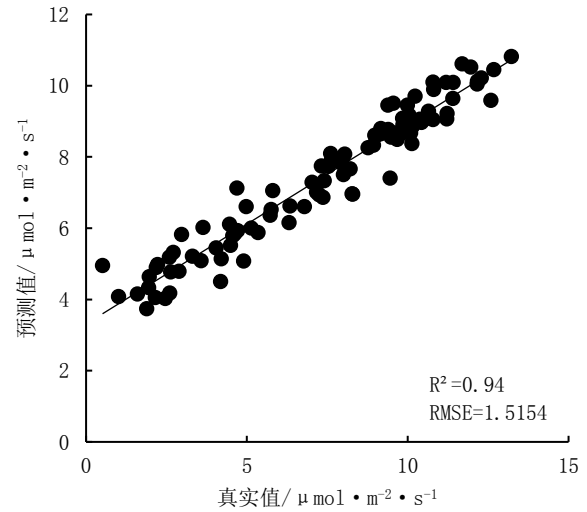


图6 随机森林反演模型

能,但是支持向量机模型在预测精度上略逊于随机森林反演模型。

## 3 讨论

本研究利用高光谱技术结合先进的机器学习算法,对分蘖洋葱叶片的净光合速率进行了定量反演。通过主成分分析(PCA)以及随机森林(Random Forest, RF)和支持向量机(Support Vector Machine, SVM)学习模型的应用,构建了高精度的高光谱反射率反演净光合速率的模型,并详细评估了模型的预测效果。

通过主成分分析对分蘖洋葱的高光谱反射率数据进行降维处理是本研究的初步步骤。这一处理不仅简化了数据结构,还提取了反映净光合速率的关键高光谱反射率特征,使得数据更加适合于后续的机器学习处理。PCA的结果表明,PC1~PC9已累计解释了数据总变异的92.9%,为构建机器学习模型提供了稳定而准确的输入特征,确保了模型训练的高效性和准确性。

在随机森林模型的构建中,该算法模型具有更高的精度,决定系数( $R^2$ )高达0.94,均方根误差(RMSE)为1.5154,显示了模型在预测分蘖洋葱叶片净光合速率上的高准确性。随机森林模型的优势在于其处理高维数据的能力,以及通过集成多个决策树来减少模型偏差,增强模型在新数据上的泛化能力。此外,随机特征选择机制也显著降低了过拟合的风险,增强了模型的稳定性和可靠性。相比之下,支持向量机模型虽然在性能指标上略逊于随机森林模型,但其决定系数为0.81,均方根误差为1.63,依然证明了其在处理具有复杂非线性关系数据时的有效性。SVM的核心在

于寻找最优的超平面,以实现数据的最好分割,这在预测精度需求较高的应用场景下尤为重要。尽管预测准确度略低,SVM因其强大的分类能力和优秀的泛化性能,仍然是处理复杂数据集的有力工具。

## 4 结 论

本研究利用高光谱技术结合机器学习算法精确分析了分蘖洋葱叶片净光合速率高光谱反演模型。通过比较随机森林(RF)和支持向量机(SVM)两种机器学习模型,发现随机森林模型凭借其卓越的预测能力和稳定性,更适宜于构建高精度的高光谱反射率反演模型。随机森林算法通过集成多个决策树,有效降低了在独立数据集上的预测误差,并显著提升了模型的预测精度和稳定性。该模型表现出的高决定系数( $R^2=0.94$ )及较低的均方根误差( $RMSE=1.5154$ ),证明了其在处理和预测分蘖洋葱净光合速率相关的高光谱数据方面具有明显优势。此外,随机森林模型在降维处理后的数据上展现出的高效性和精确性,使其成为构建复杂生物学模型的首选工具。本研究结果不仅为分蘖洋葱叶片及其他农作物的精细化管理提供了科学依据,同时为作物生长监控和产量提升提供了强大的数据支持。

### 参考文献:

- [1] 赵红星,孙雪娇,王立新,等.基于套种模式下毛葱机械收获装备的研究及发展难点与对策[J].农业与技术,2022,42(16):38-42.
- [2] 才卓伟,郑良,王丽艳,等.毛葱套种大豆高产栽培技术[J].北方园艺,2007(9):90.
- [3] 刘世欢,高志强.北方地区毛葱套种大豆高产栽培技术[J].现代农业科技,2009(18):123.
- [4] 周琬颜.中国洋葱国际竞争力研究[D].保定:河北农业大学,2019.
- [5] 刘思宇,闫文义,张慧,等.分蘖洋葱研究现状与发展前景[J].中国林副特产,2018(6):89-94.
- [6] Amanda B. Biomass and photosynthetic efficiency of *Allium* species grown in Elevated carbon dioxide levels, with differing plant densities and Harvest schemes[D]. Texas: Texas Technology University, 2009.
- [7] 蒲凯国,丁东霞,李能慧,等.玉米黄质对低温弱光下辣椒光合能力和叶黄素循环的影响[J].干旱地区农业研究,2024,42(2):177-187.
- [8] 马新国.测定光合、蒸腾速率的主从式虚拟仪器系统[D].北京:中国农业大学,2004.
- [9] 卢俊玮,田容才,花宇辉,等.光谱技术在甘蓝型油菜品质检测中的研究进展[J].激光生物学报,2020,29(4):289-294,301.
- [10] 陈鹏飞, Nicolas Tremblay, 王纪华, 等. 估测作物冠层生物量的新植被指数的研究[J]. 光谱学与光谱分析, 2010, 30(2):512-517.
- [11] 吕玮,李玉环,毛伟兵,等.基于高光谱的小麦旗叶净光合速率的遥感反演模型的比较研究[J].农业资源与环境学报,2017,34(6):582-586.
- [12] 刘聪,彭漪,方圣辉.基于高光谱数据的水稻叶片净光合速率反演[J].中国农业大学学报,2020,25(1):56-65.
- [13] 王月,徐绍棠,姚海燕,等.基于小波分析小麦旗叶净光合速率高光谱遥感反演[J].浙江农业科学,2020,61(10):1979-1984.
- [14] 董德誉,程宇坤,王睿,等.基于无人机多光谱影像反演不同生育期小麦光合参数分析[J].新疆农业科学,2023,60(6):1308-1318.
- [15] 王娣,佃袁勇,乐源,等.基于高光谱植被指数的叶片净光合速率  $P_n$  反演[J].地理与地理信息科学,2016,32(4):42-48.
- [16] Wu S, Amari SI. Conformal transformation of kernel functions: A data-dependent way to improve support vector machine classifiers[J]. Neural Processing Letters, 2002, 15: 59-67.
- [17] Breiman, Leo. Random forests[J]. Machine learning, 2001, 45: 5-32.
- [18] 李欣海.随机森林模型在分类与回归分析中的应用[J].应用昆虫学报,2013,50(4):1190-1197.
- [19] 郭宏亮,黄帅,聂楚楚,等.基于分光吸收特性的辣椒叶片叶绿素含量测量方法研究[J].东北农业科学,2021,46(5):82-86.

(责任编辑:范杰英)