

网络出版时间:2024-05-11

网络出版地址:https://link.cnki.net/urlid/61.1359.s.20240509.1041.006

## 基于植被指数特征优选的冬小麦叶片含水量估算

马宇欣,胡笑涛,王亚昆,范晓懂,彭雪莲,孙骏,陈洪

(西北农林科技大学旱区农业水土工程教育部重点实验室,陕西杨陵 712100)

**摘要:**为进一步提升利用高光谱数据快速监测叶片含水量的能力,以关中地区冬小麦为研究对象,分别获取2022和2023年孕穗期、抽穗期及灌浆期冬小麦叶片含水量,并同步监测叶片高光谱信息。通过波段组合的形式构建植被指数,并利用相关性分析进行初步筛选,再以RelieFF算法优选得到输入特征变量,然后分别利用随机森林(random forest, RF)、长短期记忆(long short-term memory, LSTM)网络和基于粒子群(particle swarm optimization, PSO)优化的反向传播神经网络(back propagation neural network, BPNN)构建冬小麦叶片含水量估算模型,并进行精度评价。结果表明,通过相关性分析与RelieFF算法结合优选变量,能够较单独通过相关分析明显提升LSTM和PSO-BPNN模型的建模精度,但对RF模型则无法优化变量。相关性分析与RelieFF结合后PSO-BPNN模型在各生育时期均取得最佳预测结果,其中孕穗期、抽穗期和灌浆期验证集 $r^2$ 分别为0.816、0.736和0.806, RMSE分别为0.546%、0.899%和1.531%, NRMSE分别为0.681%、1.195%和2.185%。由此可见,在相关分析的基础上,通过RelieFF算法优选特征变量能够提升特定模型对冬小麦叶片含水量的估算精度,其中对PSO-BPNN模型的效果最好。

**关键词:**冬小麦;叶片含水量;机器学习;变量筛选;植被指数

中图分类号:S512.1;S311

文献标识码:A

文章编号:1009-1041(2025)02-0234-11

## Estimation of Leaf Water Content of Winter Wheat Based on Vegetation Index Feature Optimization

MA Yuxin, HU Xiaotao, WANG Yakun, FAN Xiaodong, PENG Xuelian, SUN Jun, CHEN Hong

(Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** Estimation of the leaf water content (LWC) plays an important role in field irrigation management. This study aimed to estimate the LWC of winter wheat based on hyperspectral data of leaf blades, especially focusing on the effect of different variable screening methods and growth stages on the estimation model. Research data were obtained from field trials in 2022 and 2023 at the booting, heading, and grain-filling stages. Vegetation indices were constructed for each growth stage by combining the two bands. The input characteristic variables based on vegetation indices were screened by two methods: (I) the input characteristic variables were directly obtained by ranking the correlation coefficients; (II) based on the method I, the vegetation index was further screened by the RelieFF algorithm to obtain a second set of input characteristic variables. The LWC estimation models were constructed using random forest (RF), long short-term memory (LSTM) network and back propagation neural network (BPNN) based on particle swarm optimization (PSO). The best method for estimating LWC was derived by comparing the accuracy of the models. The results showed that comparing the two variable screening methods, the characteristic variables further screened by RelieFF could

收稿日期:2024-01-10 修回日期:2024-02-29

基金项目:国家自然科学基金项目(U2243235)

第一作者 E-mail: mayx0165@nwfufu.edu.cn (马宇欣)

通讯作者 E-mail: huxiaotao11@nwsuaf.edu.cn (胡笑涛)

effectively improve the accuracy of the LSTM and PSO-BPNN models, while the effect of improving the RF model is not obvious. The best model for each growth stage was established by the ReliefF screening method combined with the PSO-BPNN, at the booting stage, heading stage and grain-filling stage. The  $r^2$  of the validation set was 0.816, 0.736, and 0.806, respectively, and the RMSE was 0.546%, 0.899%, and 1.531%, respectively, and the NRMSE was 0.681%, 1.195%, and 2.185%, respectively. It was suggested that the screening method of feature variables through the ReliefF algorithm could improve its estimation accuracy in the particular model. Its combination with the PSO-BPNN model had the best application effect in the estimation of LWC in winter wheat at the growth stages.

**Keywords:** Winter wheat; Leaf moisture content; Machine learning; Variable screening; Vegetation index

叶片含水量是影响叶片光合作用和蒸腾作用的关键因素,也是诊断作物水分状况的重要指标。因此,快速、准确地获取叶片含水量对指导田间灌溉、保障作物产量具有重要意义<sup>[1]</sup>。利用光谱信息估算作物生长参数具有实时无损的优势,其中高光谱遥感技术包含丰富的连续窄波段数据,能够更准确地反映作物本身的光谱特征<sup>[2]</sup>,在一定程度上削弱了光谱中同谱异物、同物异谱造成的影响。

近年来,国内外学者利用高光谱信息对不同作物的叶片含水量进行了众多研究<sup>[3-5]</sup>。由于高光谱数据量大,波段间容易出现数据冗余,对数据进行必要降维,筛选提取特征变量是重要的研究方向。相关性分析是应用最多的变量筛选方法,但其形式单一且筛选的变量间存在多重共线性,可能影响后期建模精度<sup>[6]</sup>,因此部分学者针对多种筛选方法进行了对比研究。例如,Xu等<sup>[7]</sup>在相关性分析的基础上,通过Lasso方法去除自变量间的共线性进行特征波段选取;郭阳等<sup>[8]</sup>通过联合子区间偏最小二乘法(SiPLS)结合GA、CARS和SPA方法进行特征波长选择,结果表明SiPLS-CARS方法结果最优;崔锦涛等<sup>[9]</sup>通过六种方法筛选特征波段估算陆地棉叶片水分,发现CARS耦合SPA的方法在模型构建中效果最佳。ReliefF算法是Kononenko在1994年对Relief进行扩展得到的特征选择算法<sup>[10]</sup>,其通过比较样本间的特征差异来评估特征的重要性,研究者可通过其对回归特征变量的重要性排序实现特征优选。目前,该算法在医学、电子信息领域应用较多<sup>[11-12]</sup>,而对于田间光谱信息的特征筛选研究较少,其适用性还有待于验证。此外,对于特征变量的筛选多基于波段反射率,而对植被指数的筛选研究较少,且为了充分利用高光谱波段信息丰富

的优势,构建特定光谱波长范围内的全波段组合的植被指数成为研究热点,这使特征变量范围更大,如何在海量数据中快速寻求最佳变量组合还有待进一步研究。

随着计算机科学的不断发展,机器学习在植物叶片含水量监测中已有广泛研究。由于不同机器学习算法原理不尽相同,对于模型方法的选择是影响指标估测精度的主要原因。目前,前人的研究主要集中在随机森林、神经网络、支持向量机等方法<sup>[13-15]</sup>。同时,为解决机器学习带来的“黑箱问题”,部分研究通过群智能算法对模型进行改进。例如,张旭等<sup>[16]</sup>通过遗传算法(genetic algorithms, GA)对神经网络进行优化,实现了葡萄叶片含水量的快速估测;李铠等<sup>[17]</sup>利用粒子群算法(particle swarm optimization, PSO)对BP神经网络进行优化,构建了冬小麦冠层叶片含水量预测模型。这些结果均表明,群智能优化算法能够进一步提升模型估算效果。近年来,深度学习逐渐应用于研究中,如Zhang等<sup>[18]</sup>对比了传统机器学习模型和深度学习模型在油菜叶片含水量估算中的潜力,结果表明CNN-LSTM模型估算效果最佳,测试集 $r^2$ 达0.814。综合来看,随着模型构建方法的不断增多,在前人研究的基础上,如何择优选择估算模型仍是热点问题。

上述对于叶片含水量的研究均基于特定生育时期或全生育期进行,而不同生育时期光谱特征和叶片含水量间的关系有所差异,研究特定生育时期具有局限性,全生育期则可能因为同谱异物、同物异谱等问题影响估算精度,因此分生育时期对叶片含水量的研究具有重要意义。针对以上问题,本研究分别对冬小麦孕穗期、抽穗期和灌浆期3个主要生育时期进行两年试验数据获取,计算各生育时期全波段植被指数,并通过相关性分析

对植被指数进行初筛,同时应用 ReliefF 算法优选特征变量,由随机森林(random forest, RF)、长短期记忆(long short-term memory, LSTM)网络及基于粒子群优化的反向传播神经网络(back propagation neural network, BPNN)构建不同生育时期叶片含水量估算模型,通过对比分析不同变量筛选方法与建模方法组合所获的建模精度,探究冬小麦不同生育时期最优叶片含水量估算模型,以期应用特征筛选方法提升模型精度研究提供参考。

## 1 材料与方法

### 1.1 研究区概况

本研究于西北农林科技大学灌溉试验站进行。该站位于关中平原中部(34°20'N, 108°24'E),属暖温带半湿润气候区,多年平均气温 12.9℃,年平均降水量约 630 mm,降水主要集中于 7—9 月。试验采用随机区组设计,按照灌溉计划湿润层深度,设置 4 个深蓄灌溉水平,分别为 W1 (1.2 m 计划湿润层)、W2 (1.4 m 计划湿润层)、W3 (1.6 m 计划湿润层)、W4 (1.8 m 计划湿润层)及 CK (0.8 m 计划湿润层,常规灌溉),其中 CK 设置 3 个重复,其余处理设置 6 个重复,共 27 个试验小区,小区面积为 12.8 m<sup>2</sup>,各小区间隔 0.8 m,并在试验地四周设置 1.0 m 的保护行。供试小麦品种为小偃 22,两季小麦播种时间分别为 2021 年 10 月 20 日和 2022 年 10 月 16 日。

### 1.2 数据获取

本研究 2022 年数据获取时间分别为 4 月 10 日(孕穗期)、4 月 23 日(抽穗期)和 5 月 16 日(灌浆期),2023 年分别为 4 月 8 日(孕穗期)、4 月 15 日(灌浆期)、4 月 19 日(抽穗期)、4 月 26 日(抽穗期)和 5 月 13 日(灌浆期)、5 月 24 日(灌浆期)。

### 1.2.1 叶片高光谱数据获取与预处理

采用便携式手持光谱仪 PolyPen RP410 NIR 进行叶片光谱反射率测定。该仪器内集成一个内部光源(氙气白炽灯),光束以固定入射角 35°照射待测样本,可快速得出所测叶片光谱反射率,测量波段范围为 629~1 060 nm,重采样间隔为 1 nm。每次使用前用校准卡进行白色反射标准的校准,并开启 GPS 定位,设置读数扫描次数为三,即所得数据为三次扫描的平均值。每个小区中选择 5 株具有长势代表性的冬小麦,测量其上、中、下三片叶子的叶片高光谱数据,每个小区共获取 15 条叶片光谱曲线。将获取的数据导入 Spectrapen 光谱分析软件中进行原始光谱反射率导出,并计算 15 条光谱曲线的平均值代表整个小区的叶片光谱数据。

### 1.2.2 叶片含水量数据获取

于每个小区中随机选取长势均匀的 10 株小麦,摘取茎秆所有叶片迅速称鲜重,并将叶片装袋,放入 105℃烘箱中杀青 1 h,恒温 75℃烘干至恒重后,称量干重。叶片含水量=[(鲜重-干重)/鲜重]×100%。

### 1.3 植被指数选取与计算

植被指数能够综合光谱信息,增强植被信息<sup>[19-20]</sup>。结合前人的研究成果,本研究通过两波段组合的计算方式,选取 4 种常用的植被指数组合形式评估其在冬小麦叶片含水量估算中的应用潜力(表 1)。

### 1.4 变量筛选

选用皮尔逊相关系数法和 ReliefF 两种方法对模型输入变量进行筛选。筛选思路:(1)对不同生育时期构建的全波段植被指数与对应叶片含水量进行相关性分析,并对相关系数进行排序,筛选前 5 个植被指数作为第一组输入变量;(2)采用

表 1 植被指数及其计算公式

Table 1 Vegetation indices and their calculation formulas

植被指数 Vegetation index	计算公式 Computational formula	参考文献 Reference
比值植被指数 Ratio vegetation index, RVI	$RVI=R_i/R_j$	[21]
两波段增强植被指数 Two bands enhanced vegetation index, EVI2	$EVI2=2.5 \times (R_i - R_j) / (R_i + 2.4 \times R_j + 1)$	[22]
归一化红边植被指数 Normalized difference red edge index, NDRE	$NDRE=(R_i - R_j) / (R_i + R_j)$	[23]
差值植被指数 Difference vegetation index, DVI	$DVI=R_i - R_j$	[24]

ReliefF 算法对各生育时期中各类植被指数相关系数排序前 100 的植被指数再次筛选,选择特征权值排序前 5 的植被指数作为第二组输入变量。

在这两种特征变量筛选方法中,皮尔逊相关系数法简单快捷,在当前研究领域应用广泛。其通过计算不同变量与研究指标间的相关系数,来度量两者间的相关性,相关系数绝对值越接近于 1,表明二者相关性越好。ReliefF 算法作为 Relief 算法的进阶,能够实现多分类,降低样本特征维度<sup>[25]</sup>。主要原理是每次从训练集中随机抽取一个样本,分别计算与样本同类别和不同类别中近邻样本的距离,通过不断迭代计算,比较类间距和类内距的大小以更新特征权值大小,从而进行特征选择。

### 1.5 模型构建与验证

基于田间试验于每个生育时期分别获取 81 个叶片光谱和冬小麦叶片含水量样本组合,按照 2:1 的比例划分建模集(54 个)与验证集(27 个)。根据前人的研究结果<sup>[26-28]</sup>,本研究选择随机森林(RF)、长短期记忆(LSTM)神经网络和粒子群算法优化反向传播神经网络(PSO-BPNN),使用 MATLAB R2022a 软件构建冬小麦水分指标估算模型。

随机森林是一种基于集成学习的机器学习算法,应用 Bagging 方法在训练集 T 中有放回地随机抽取  $N$  个样本,通过构建多个决策树完成分类和回归任务<sup>[29]</sup>;其训练速度快且可解释性强,能够评估不同特征间的重要程度。长短期记忆神经网络是一种特殊的递归神经网络(recurrent neural network, RNN),能够有效解决解决 RNN 结构中的梯度消失和梯度爆炸问题<sup>[30]</sup>;其通过门限管理存储单元信息,分别包括遗忘门、输入门及输出门,能更好捕捉序列数据中的长期依赖关系<sup>[18]</sup>。反向传播神经网络(BPNN)是人工神经网络中应用广泛的多层前馈神经网络,通过对误差逆传播进行不断修正,使实际输出正确率不断提升;其具有较强的非线性映射能力、学习和自适应能力,适用于内部机制复杂的问题<sup>[31]</sup>。由于 BPNN 模型的网络权值初始化是随机的,模型的稳定性还需进一步优化<sup>[32]</sup>,因而本研究选取粒子群算法,从随机解出发,通过迭代寻求每个粒子最优解,用适应度来评价解的品质,确定 BPNN 模型的最佳权值和阈值,以达到优化目的<sup>[33]</sup>。

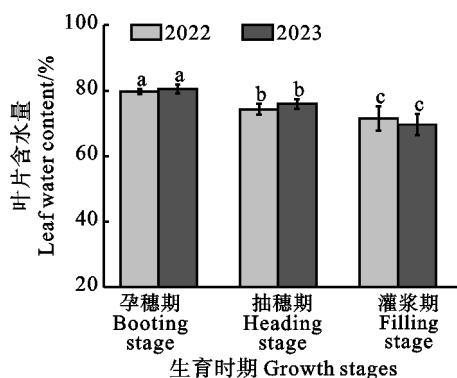
模型评价指标选用决定系数(coefficient of

determination,  $r^2$ )、均方根误差(root mean squared error, RMSE)以及标准均方根误差(normalized root mean squared error, NRMSE)。其中,决定系数取值越接近 1,表明预测模型的拟合效果越好;RMSE 值越小,表明预测值与实测值一致性越好;NRMSE 值越低,表明预测模型精度越高。

## 2 结果与分析

### 2.1 不同生育时期冬小麦叶片含水量分布及光谱响应特征

2022 年和 2023 年冬小麦叶片含水量在各生育期的变化趋势一致(图 1),均表现为从孕穗期逐渐递减,灌浆期达到最小值,且经方差分析不同生育时期均具有显著性差异( $P < 0.05$ )。两年数据相比,2022 年孕穗期、抽穗期低于 2023 年,灌浆期由于 2022 年对冬小麦进行了补灌,叶片含水量高于 2023 年。



图柱上的字母表示不同时期差异在 0.05 水平上显著。

Different letters above the columns indicate significant differences at 0.05 level.

图 1 不同生育时期冬小麦叶片含水量变化

Fig. 1 Changes of leaf water content of winter wheat at different growth stages

从冬小麦叶片原始光谱反射率变化曲线(图 2)看,2022 年与 2023 年相比整体略低,可能与气候条件、种植时间等因素有关。但两年内光谱反射率的生育期总体变化趋势基本一致,表明冬小麦叶片光谱变化在不同生育时期具有一定规律性。其中,抽穗期反射率在波段 750~1 020 nm 明显高于孕穗期和灌浆期;孕穗期反射率在波段 629~690 nm 明显低于抽穗期和灌浆期;在波段 690~750 nm 三个生育时期反射率变化曲线相似。

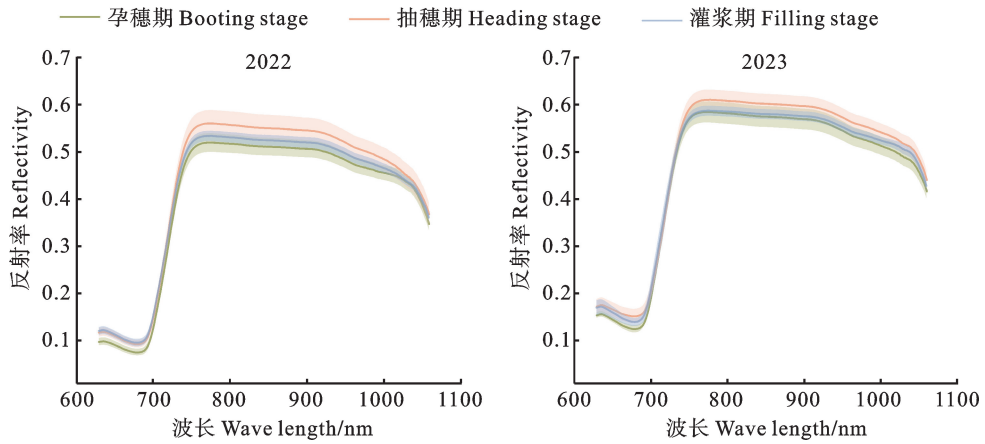


图 2 不同生育时期冬小麦叶片光谱响应曲线

Fig. 2 Spectral response curves of winter wheat leaves at different growth stages

### 2.2 不同生育时期植被指数筛选

#### 2.2.1 基于相关性分析筛选植被指数

对 4 种植被指数分别在 629~1 060 nm 范围内进行全波段组合,计算其与各生育时期叶片含水量的相关性并绘制相关性可视化热图(图 3),筛选植被指数相关性绝对值排序前 5 的组合用于模型构建,结果如表 2 所示。

从图 3 可知,植被指数在不同生育时期与叶片含水量的相关性分布具有一定的相似性。叶片含水量的强相关区域主要集中于波段 930~1 000 nm 与 750~950 nm 的组合范围,其中灌浆期与其他两个生育时期在相同波段顺序组合下相关性系数符号相反,可能是因为叶片光谱主要取决于叶片形态变化,而灌浆期补灌对叶片形态影响较小,相比于叶片水分含量的变化幅度较小,因此呈现相反的情况。除此之外,抽穗期在波段 630~1 060 nm 与 630~730 nm 的组合范围也具有较强相关性。不同植被指数在相同生育时期与叶片含水量相关性变化的区域差异不明显,主要集中于相关系数值的大小差异。由表 2 可知,孕穗期

和抽穗期植被指数 RVI 和 NDRE 表现均较佳,相关系数最大绝对值分别为 0.724、0.736。灌浆期植被指数 EVI2 最佳,相关系数绝对值为 0.798。

#### 2.2.2 基于 ReliefF 筛选植被指数

基于相关性分析筛选各植被指数相关性排序前 100 的变量作为 ReliefF 筛选数据集,即每个生育时期得到 400 个变量,然后通过特征权值判断其重要性,选择排序前 5 的植被指数作为模型输入变量。ReliefF 方法对输入植被指数的特征权值由图 4 所示,表 3 分别列出特征权值排序前 5 的植被指数。两种方法筛出的植被指数不同,且相比于相关性分析,ReliefF 并未筛选相关性最佳的植被指数作为输入变量。

### 2.3 叶片含水量估算模型构建及模型精度检验

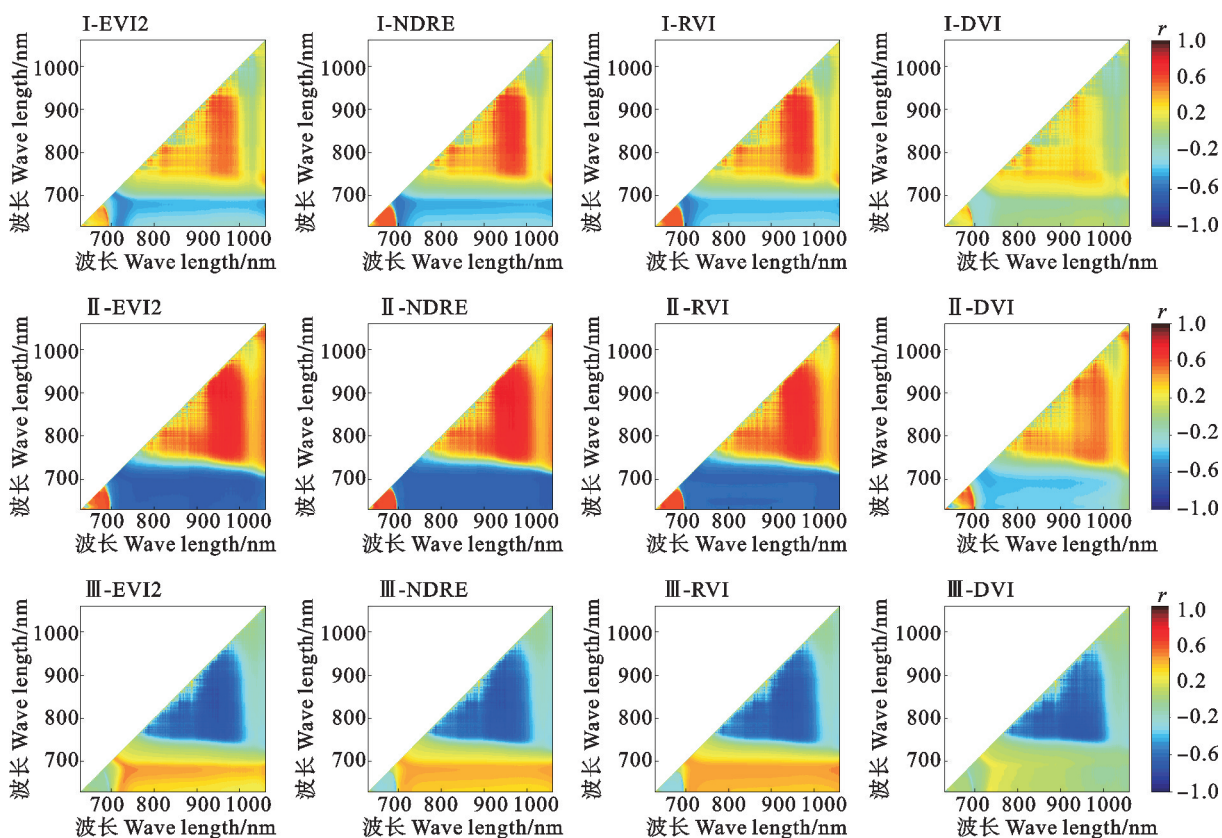
#### 2.3.1 基于 RF 模型估算冬小麦叶片含水量

基于两类输入变量分别构建 RF 模型。从不同生育时期来看,孕穗期和灌浆期预测精度相近,且均高于抽穗期(图 5)。通过 ReliefF 算法筛选最佳变量的方法对 RF 模型仅在灌浆期对建模和

表 2 不同生育时期相关性排序前 5 的植被指数

Table 2 Vegetation index of the top 5 correlation ranking at different growth stages

孕穗期 Booting stage			抽穗期 Heading stage			灌浆期 Filling stage		
植被指数 Vegetation index	r	波段组合 Band combination	植被指数 Vegetation index	r	波段组合 Band combination	植被指数 Vegetation index	r	波段组合 Band combination
RVI	0.724	926 nm, 955 nm	RVI	0.736	907 nm, 939 nm	EVI2	0.798	838 nm, 948 nm
NDRE	0.723	926 nm, 955 nm	NDRE	0.736	907 nm, 939 nm	EVI2	0.797	856 nm, 948 nm
RVI	0.723	926 nm, 954 nm	RVI	0.733	898 nm, 943 nm	EVI2	0.797	867 nm, 936 nm
NDRE	0.723	926 nm, 954 nm	NDRE	0.733	898 nm, 943 nm	EVI2	0.796	868 nm, 936 nm
RVI	0.720	926 nm, 964 nm	RVI	0.732	897 nm, 939 nm	EVI2	0.795	837 nm, 948 nm



I、II、III 分别代表孕穗期、抽穗期和灌浆期。

I, II and III represent the booting stage, heading stage, and filling stage, respectively.

图 3 不同生育时期植被指数与叶片含水量的相关矩阵图

Fig. 3 Correlation matrix between vegetation index and leaf water content at different growth stages

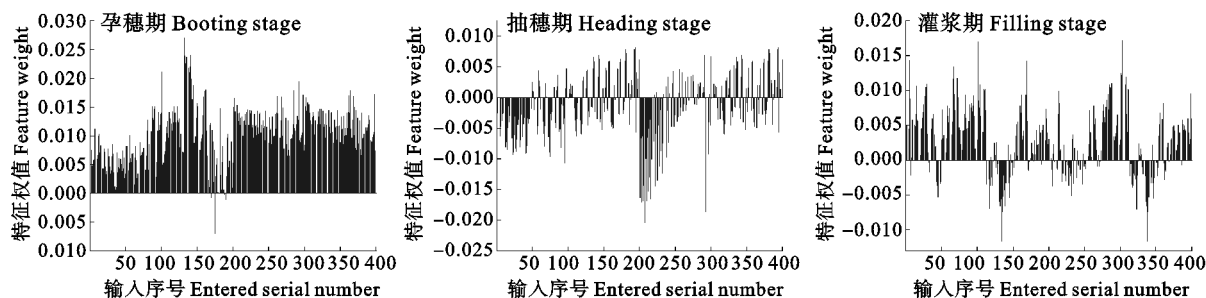


图 4 不同小麦生育时期筛选植被指数的特征权值

Fig. 4 Feature weights of screened vegetation indices at different wheat growth stages

表 3 不同生育时期特征权值排序前 5 的植被指数

Table 3 Vegetation index of the top 5 in the ranking of feature weights at different growth stages

孕穗期 Booting stage			抽穗期 Heading stage			灌浆期 Filling stage		
植被指数 Vegetation index	$ r $	波段组合 Band combination	植被指数 Vegetation index	$ r $	波段组合 Band combination	植被指数 Vegetation index	$ r $	波段组合 Band combination
NDRE	0.703	880 nm, 971 nm	DVI	0.623	923 nm, 934 nm	NDRE	0.776	856 nm, 916 nm
NDRE	0.703	881 nm, 973 nm	NDRE	0.725	907 nm, 960 nm	RVI	0.776	856 nm, 916 nm
NDRE	0.703	884 nm, 972 nm	RVI	0.725	907 nm, 960 nm	EVI2	0.781	867 nm, 929 nm
NDRE	0.702	880 nm, 972 nm	NDRE	0.724	907 nm, 959 nm	NDRE	0.765	884 nm, 936 nm
NDRE	0.699	881 nm, 971 nm	RVI	0.724	907 nm, 959 nm	EVI2	0.791	831 nm, 948 nm

验证精度均有提升效果,且提升幅度较小,其中建模集  $r^2$  由 0.803 提升至 0.805,验证集由  $r^2$  由 0.748 提升至 0.761, RMSE 和 NRMSE 均小幅下降;孕穗期中基于 ReliefF 算法筛选的变量仅在验证集中精度高于相关性分析结果,在建模集未表现出精度提升效果,建模集和验证集的最佳估算结果中  $r^2$  分别为 0.769 和 0.773, RMSE 分别为 0.619% 和 0.765%, NRMSE 分别为 0.771% 和 0.749%;抽穗期建模结果则与孕穗期相反,ReliefF 仅在建模集中有精度提升效果,其中建模集和验证集的最佳估算结果中  $r^2$  分别为 0.734 和 0.684, RMSE 分别为 0.854% 和 0.982%, NRMSE 分别为 1.133% 和 1.305%。总体来看,ReliefF 算法对 RF 模型估算叶片含水量精度的提升效果不明显,各生育时期模型精度变化无规律。

2.3.2 基于 LSTM 模型估算冬小麦叶片含水量

基于两类输入变量分别构建 LSTM 模型。从不同生育时期来看,其规律变化与 RF 模型结果相同,仍体现出孕穗期和灌浆期精度高于抽穗期精度(图 6)。但通过 ReliefF 算法筛选最佳变量的方法对 LSTM 模型精度在每个生育时期均有提升效果,其中孕穗期建模集和验证集  $r^2$  分别提升 0.046 和 0.115, RMSE 分别下降 0.052% 和

0.129%, NRMSE 分别下降 0.064% 和 0.162%;抽穗期建模集和验证集  $r^2$  分别提升 0.111 和 0.093, RMSE 分别下降 0.169% 和 0.124%, NRMSE 分别下降 0.224% 和 0.164%;灌浆期建模集和验证集  $r^2$  分别提升 0.050 和 0.047, RMSE 分别下降 0.145% 和 0.205%, NRMSE 分别下降 0.206% 和 0.293%。由此可得,通过 ReliefF 算法对特定植被指数筛选有助于提高 LSTM 模型的估算精度。

2.3.3 基于 PSO-BPNN 模型估算冬小麦叶片含水量

基于两类输入变量分别构建 PSO-BPNN 模型。通过 ReliefF 算法筛选的变量组合在 PSO-BPNN 模型中建模集和验证集总体结果均优于基于相关性分析筛选结果,表明 ReliefF 算法对特定植被指数筛选有助于提高 PSO-BPNN 模型的估算精度(图 7)。其中,孕穗期建模集和验证集  $r^2$  分别为 0.783 和 0.816, RMSE 分别为 0.585% 和 0.546%, NRMSE 分别为 0.729% 和 0.681%;抽穗期建模集和验证集  $r^2$  分别为 0.768 和 0.736, RMSE 分别为 0.803% 和 0.899%, NRMSE 分别为 1.065% 和 1.195%;灌浆期建模集和验证集  $r^2$  分别为 0.813 和 0.806, RMSE 分别为 1.526% 和 1.531%, NRMSE 分别为 2.169% 和 2.185%。

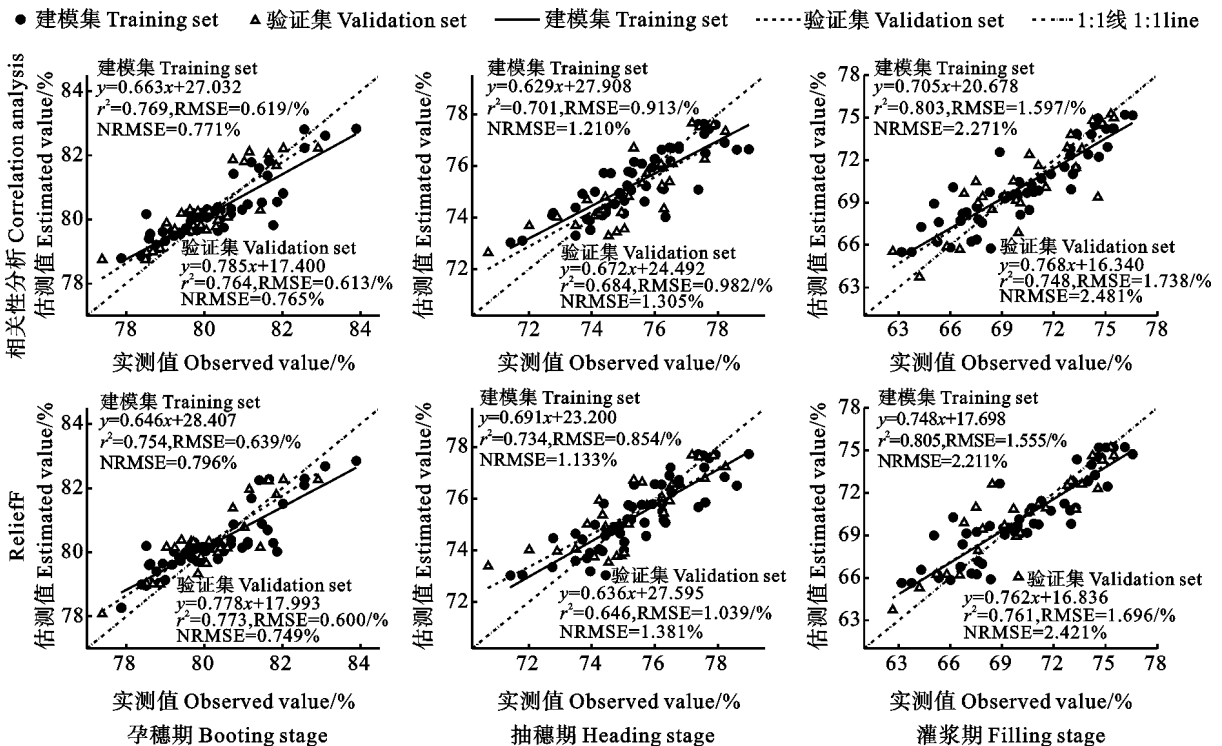


图 5 基于 RF 的冬小麦叶片含水量估算模型

Fig. 5 Estimation model of water content in winter wheat leaves based on RF

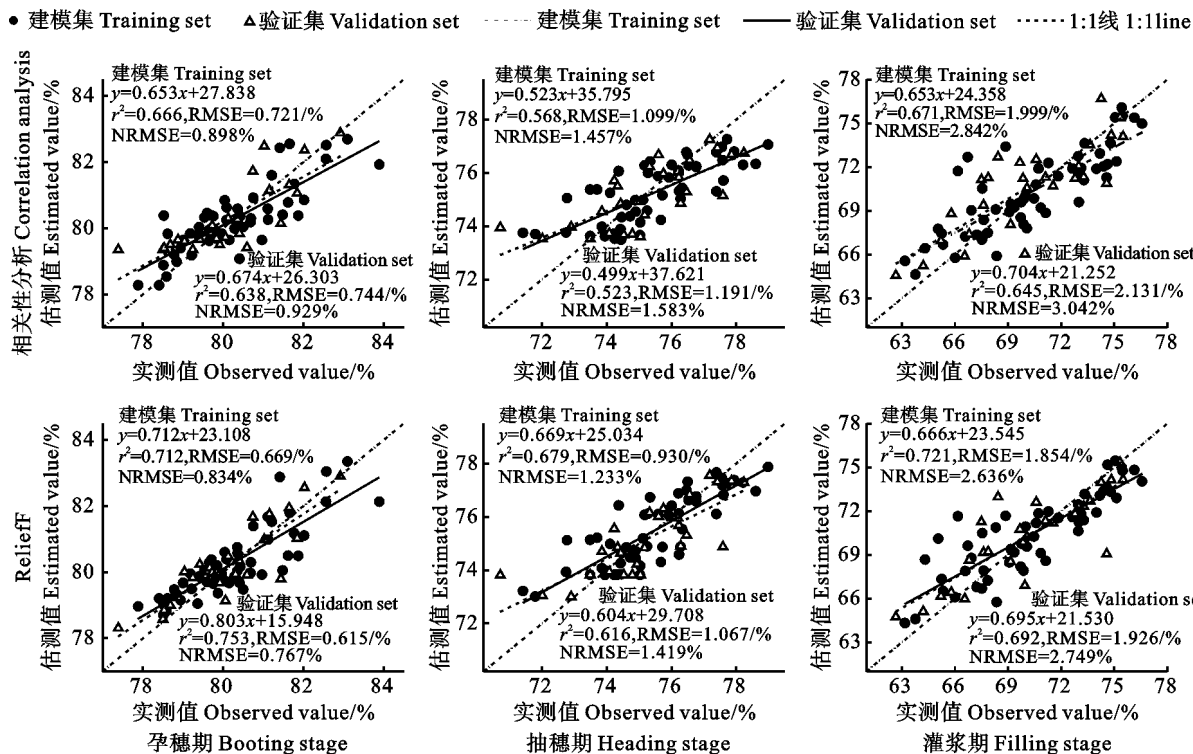


图 6 基于 LSTM 的冬小麦叶片含水量估算模型

Fig. 6 Estimation model of water content in winter wheat leaves based on LSTM

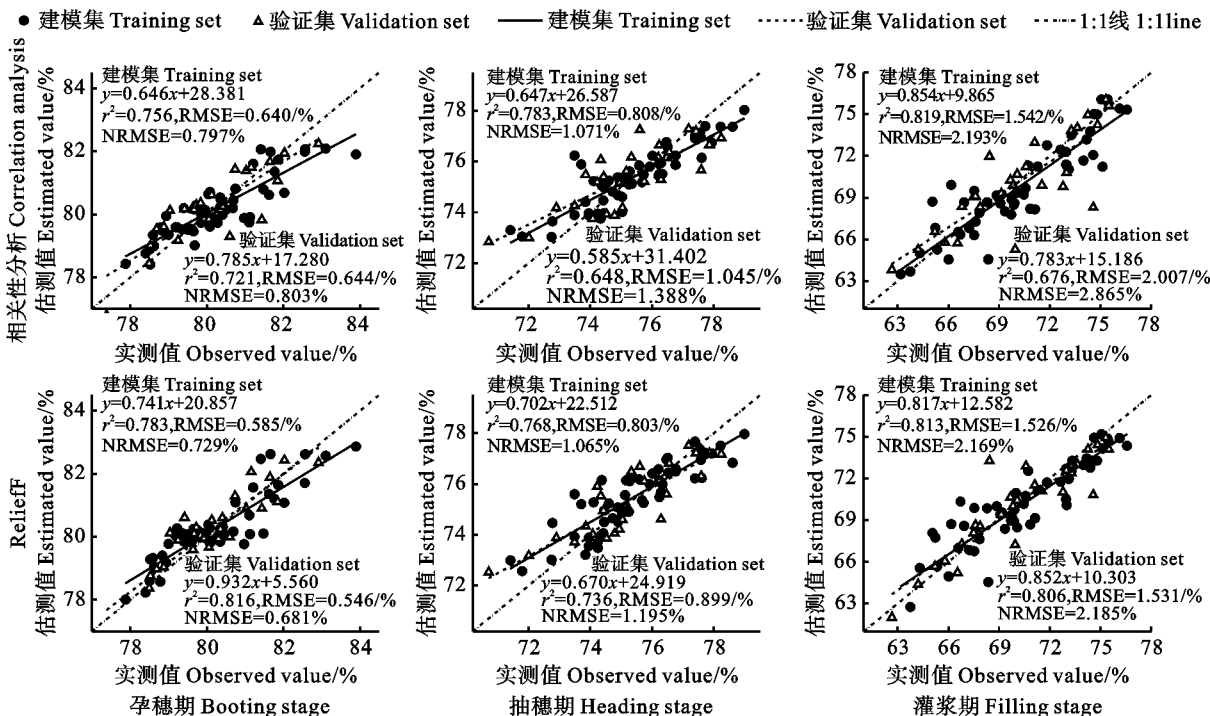


图 7 基于 PSO-BPNN 的冬小麦叶片含水量估算模型

Fig. 7 Estimation model of water content in winter wheat leaves based on PSO-BPNN

从三种模型的估算结果来看,尽管 Relieff 优选的变量组合能够进一步提升 LSTM 模型的估算精度,但其结果总体仍低于 RF 和 PSO-

BPNN 模型。其中,RF 和 PSO-BPNN 模型在基于相关性分析筛选的变量集中估算结果相当,但在基于 Relieff 优选的变量集中 PSO-BPNN 模

型的总体建模精度更优于 RF 模型,最终各生育时期均在 ReliefF 优选变量组合结合 PSO-BPNN 模型中取得最佳估算结果,孕穗期、抽穗期和灌浆期的验证集  $r^2$  分别为 0.816、0.736 和 0.806。

### 3 讨论

本研究基于叶片高光谱信息计算 4 个全波段植被指数,由相关性分析和 ReliefF 算法分别筛选出 5 个输入变量,分别构建冬小麦叶片含水量在孕穗期、抽穗期及灌浆期的 RF、LSTM 和 PSO-BPNN 模型,发现不同输入变量和模型的组合的估算效果存在差异。

基于全波段组合构建的 4 个植被指数,在相同生育时期与冬小麦叶片含水量间的相关性大小变化趋势总体表现出相似性,这与唐子竣等<sup>[34]</sup>对冬油菜氮素营养诊断的研究结果一致,表明研究波段组合能够确定植被指数与对应指标间的最优组合范围,充分应用了高光谱的优势。本研究的波段范围集中于 629~1 060 nm,相比于前人利用光谱信息对水分的研究缺少了部分红外波段。但本研究重点在于植被指数的筛选,且相比于前人的研究成果,现有研究波段范围包括了植被指数主要敏感波段组合。Kong 等<sup>[35]</sup>对冬小麦叶片含水量的估算研究中构建了波段组合的 NDW-SI,发现 770~850 nm 和 1 000~1 110 nm 的组合与叶片含水量相关性较好;王敬湧等<sup>[36]</sup>对玉米叶片含水量的研究中,植被指数在 800~1 000 nm 波段范围组合相关性结果较好。本研究三个生育时期内,波段 930~1 000 nm 与 750~950 nm 的组合范围均表现出与叶片含水量较强的相关性,与上述研究最佳组合范围相似,后续研究中可重点明确光谱研究范围,避免过多数据处理。

不同模型对冬小麦叶片含水量的估算呈现不同结果。总体来看,RF 和 PSO-BPNN 模型的估算效果较好,其中 RF 模型应用相对成熟,前人多项研究均表明其在小样本估算模型中具有优势<sup>[37-38]</sup>;BPNN 模型虽具有较强的非线性映射能力,但其通过梯度下降算法调整权阈值,模型容易陷入局部最优解。PSO 算法能够优化 BPNN 模型寻求最佳权阈值的过程,从而提高收敛速度,获取更高预测精度,这与李铠等<sup>[17]</sup>研究结果一致。基于 LSTM 构建的模型精度相比其他模型明显较低,可能是因为 LSTM 模型的优点在于能够更

好捕捉序列数据中的依赖关系,更适合训练大型数据集<sup>[18]</sup>。对比两种特征变量筛选方法,相关性分析法通过两变量间的线性相关性筛选变量,在一定程度上忽视了植被指数和研究指标间的非线性相关性<sup>[39]</sup>,且可能导致输入变量间存在多重共线性,容易造成模型过拟合。从建模结果来看,基于 ReliefF 算法筛选的特征变量能够明显提升 LSTM 和 PSO-BPNN 的估算精度,对 RF 模型效果则不明显,可能是因为 RF 模型的自身优势在于对于数据的降维,在一定程度上防止数据过拟合,具有较强的抗干扰能力。

另外,分析本研究在不同生育时期的模型结果发现,在相同模型估算中孕穗期和灌浆期对叶片含水量的估算精度均高于抽穗期,可能是由于抽穗期冬小麦叶片已完全发育,叶片中水分含量对局部叶片形态影响较小,难以捕捉与水分相关的光谱信息。叶片光谱信息能够直接反映不受土壤等背景因素干扰的叶片光谱信息,更有利于探索植株内部水分状况与光谱间的关系。本研究仅使用单一品种进行研究,模型结果对其他品种的适用性还需进一步试验。在实际应用中,叶片光谱数据获取方式较局限,且存在以点带面的问题,后续研究还需结合无人机、卫星等先进快捷的光谱获取方式进行探讨分析,实现叶片光谱与其他光谱间的尺度转换问题。

### 4 结论

本研究通过获取 3 个生育时期的冬小麦叶片高光谱信息,利用相关性分析和 ReliefF 算法进行两波段植被指数优选,构建基于 RF、LSTM 和 PSO-BPNN 模型的冬小麦叶片含水量估算模型。结果表明,通过 ReliefF 算法优选植被指数虽无法提升 RF 模型的估算精度,但对于 LSTM 和 PSO-BPNN 模型具有明显提升效果。孕穗期、抽穗期和灌浆期均在 ReliefF 优选植被指数结合 PSO-BPNN 模型中获取叶片含水量估算最佳精度,建模集和验证集  $r^2$  分别为 0.783 和 0.816、0.768 和 0.736、0.813 和 0.806。

#### 参考文献:

- [1] 靳宁,张东彦,李振海,等.基于多光谱卫星模拟波段反射率的冬小麦水分状况评估[J].农业机械学报,2020,51(11):244.  
JIN N,ZHANG D Y,LI Z H,et al. Evaluation of water status of winter wheat based on simulated reflectance of multispectral satellites [J]. *Transactions of the Chinese Society for Ag-*

- gricultural Machinery*, 2020, 51(11):244.
- [2] 刘建刚, 赵春江, 杨贵军, 等. 无人机遥感解析田间作物表型信息研究进展[J]. 农业工程学报, 2016, 32(24):98.  
LIU J G, ZHAO C J, YANG G J, *et al.* Review of field-based phenotyping by unmanned aerial vehicle remote sensing platform [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(24):98.
- [3] ZHAO J, LI H, CHEN C, *et al.* Detection of water content in lettuce canopies based on hyperspectral imaging technology under outdoor conditions [J]. *Agriculture*, 2022, 12(11):1796.
- [4] NDLOVU H S, ODINDI J, SIBANDA M, *et al.* A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)-based proximal and remotely sensed data [J]. *Remote Sensing*, 2021, 13(20):4091.
- [5] JUNTILA S, HÖLTTÄ T, SAARINEN N, *et al.* Close-range hyperspectral spectroscopy reveals leaf water content dynamics [J]. *Remote Sensing of Environment*, 2022, 277:113071.
- [6] 程介虹, 陈争光, 衣淑娟. 最小相关系数的多元校正波长选择算法[J]. 光谱学与光谱分析, 2022, 42(3):719.  
CHENG J H, CHEN Z G, YI S J. Wavelength selection algorithm based on minimum correlation coefficient for multivariate calibration [J]. *Spectroscopy and Spectral Analysis*, 2022, 42(3):719.
- [7] XU Z H, LI B, YU H, *et al.* Changing relationships between water content and spectral features in moso bamboo leaves under *Pantana phyllostachysae* Chao stress [J]. *Forests*, 2023, 14(4):702.
- [8] 郭阳, 郭俊先, 史勇, 等. SiPLS-CARS 与 GA-ELM 对哈密瓜冠层叶片含水率的反演估测[J]. 光谱学与光谱分析, 2022, 42(8):2565.  
GUO Y, GUO J X, SHI Y, *et al.* Estimation of leaf moisture content in cantaloupe canopy based on SiPLS-CARS and GA-ELM [J]. *Spectroscopy and Spectral Analysis*, 2022, 42(8):2565.
- [9] 崔锦涛, 买买提·沙吾提. 基于特征波段选择和机器学习的陆地棉叶片水分估算[J]. 干旱区地理, 2023, 46(11):1836.  
CUI J T, MAMAT SAWUT. Estimation of leaf water content in upland cotton based on feature band selection and machine learning [J]. *Arid Land Geography*, 2023, 46(11):1836.
- [10] KONONENKO I. Estimating attributes: Analysis and extensions of Relief [M]//Machine Learning; ECML-94. Berlin, Heidelberg; Springer Berlin Heidelberg, 1994:171.
- [11] AL-HADDAD L A, ALAWEE W H, BASEM A. Advancing task recognition towards artificial limbs control with ReliefF-based deep neural network extreme learning [J]. *Computers in Biology and Medicine*, 2024, 169:107894.
- [12] 薛露宇, 宋燕. 一种具有缺失数据的无监督 ReliefF 特征选择算法[J]. 小型微型计算机系统, 2023, 44(7):1442.  
XUE L Y, SONG Y. Unsupervised ReliefF feature selection algorithm with missing data [J]. *Journal of Chinese Computer Systems*, 2023, 44(7):1442.
- [13] 闫成川, 曲延英, 陈全家, 等. 基于无人机多光谱影像的棉花 SPAD 值及叶片含水量估测[J]. 农业工程学报, 2023, 39(2):61.  
YAN C C, QU Y Y, CHEN Q J, *et al.* Estimation of cotton SPAD value and leaf water content based on UAV multispectral images [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(2):61.
- [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] 尼加提·卡斯木, 张志从, 吾木提·艾山江, 等. 基于三波段光谱指数的春小麦叶片水分含量估算[J/OL]. 麦类作物学报, <http://kns.cnki.net/kcms/detail/61.1359.s.20231114.1442.012.html>.  
NIJAT K, ZHANG Z C, UMUT H, *et al.* Estimation of leaf water content of spring wheat based on 3D spectral index [J/OL]. *Journal of Triticeae Crops*. <http://kns.cnki.net/kcms/detail/61.1359.s.20231114.1442.012.html>.
- [16] 张旭, 高何璇, 高晓阳, 等. 基于低空遥感与 GA-BP 神经网络的葡萄叶片含水量估算研究[J]. 林业机械与木工设备, 2022, 50(6):69.  
ZHANG X, GAO H X, GAO X Y, *et al.* Research on estimation of water content in grape leaves based on low-altitude remote sensing and GA-BP neural network [J]. *Forestry Machinery & Woodworking Equipment*, 2022, 50(6):69.
- [17] 李铠, 常庆瑞, 陈倩, 等. 基于连续小波变换耦合 CARS 算法的冬小麦冠层叶片含水量估算[J]. 麦类作物学报, 2023, 43(2):251.  
LI K, CHANG Q R, CHEN Q, *et al.* Estimation of water content in canopy leaf of winter wheat based on continuous wavelet transform coupled CARS algorithm [J]. *Journal of Triticeae Crops*, 2023, 43(2):251.
- [18] ZHANG C, LI C, HE M Y, *et al.* Leaf water content determination of oilseed rape using near-infrared hyperspectral imaging with deep learning regression methods [J]. *Infrared Physics & Technology*, 2023, 134:104921.
- [19] MAIMAITIJIANG 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.
- [20] PENG X L, CHEN D Y, ZHOU Z J, *et al.* Prediction of the nitrogen, phosphorus and potassium contents in grape leaves at different growth stages based on UAV multispectral remote sensing [J]. *Remote Sensing*, 2022, 14(11):2659.
- [21] SCHUERGER A C, CAPELLE G A, DI BENEDETTO J A, *et al.* Comparison of two hyperspectral imaging and two laser-induced fluorescence instruments for the detection of zinc stress and chlorophyll concentration in Bahia grass (*Paspal-*

- um notatum Flugge.) [J]. *Remote Sensing of Environment*, 2003, 84(4): 572.
- [22] JIANG Z Y, HUETE A R, DIDAN K, *et al.* Development of a two-band enhanced vegetation index without a blue band [J]. *Remote Sensing of Environment*, 2008, 112(10): 3833.
- [23] FITZGERALD G, RODRIGUEZ D, O'LEARY G. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI) [J]. *Field Crops Research*, 2010, 116(3): 318.
- [24] BROGE N H, MORTENSEN J V. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data [J]. *Remote Sensing of Environment*, 2002, 81(1): 45.
- [25] 项颂阳, 许章华, 张艺伟, 等. 高光谱图像分类的 ReliefF-RFE 特征选择算法构建与应用[J]. *光谱学与光谱分析*, 2022, 42(10): 3284.
- XIANG S Y, XU Z H, ZHANG Y W, *et al.* Construction and application of ReliefF-RFE feature selection algorithm for hyperspectral image classification [J]. *Spectroscopy and Spectral Analysis*, 2022, 42(10): 3284.
- [26] 井宇航, 郭燕, 张会芳, 等. 无人机飞行高度对冬小麦植株氮积累量预测模型的影响[J]. *河南农业科学*, 2022, 51(2): 147.
- JING Y H, GUO Y, ZHANG H F, *et al.* Effects of UAV flight height on prediction model of plant nitrogen accumulation in winter wheat [J]. *Journal of Henan Agricultural Sciences*, 2022, 51(2): 147.
- [27] 李长春, 施锦锦, 马春艳, 等. 基于小波变换和分数阶微分的冬小麦叶绿素含量估算[J]. *农业机械学报*, 2021, 52(8): 172.
- LI C C, SHI J J, MA C Y, *et al.* Estimation of chlorophyll content in winter wheat based on wavelet transform and fractional differential [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(8): 172.
- [28] HUANG X, GUAN H D, BO L Y, *et al.* Hyperspectral proximal sensing of leaf chlorophyll content of spring maize based on a hybrid of physically based modelling and ensemble stacking [J]. *Computers and Electronics in Agriculture*, 2023, 208: 107745.
- [29] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. *信息技术*, 2018, 12(1): 50.
- WANG Y S, XIA S T. A survey of random forests algorithms [J]. *Information and Communications Technologies*, 2018, 12(1): 50.
- [30] 赵东东, 赵雅丽, 赵秉强, 等. 基于 PCA-LSTM 小麦叶片水分检测研究[J]. *中国农机化学报*, 2019, 40(3): 155.
- ZHAO D D, ZHAO Y L, ZHAO B Q, *et al.* Study on wheat leaf moisture detection based on PCA-LSTM [J]. *Journal of Chinese Agricultural Mechanization*, 2019, 40(3): 155.
- [31] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. *计算机工程与应用*, 2021, 57(11): 59.
- ZHANG C, GUO Y, LI M. Review of development and application of artificial neural network models [J]. *Computer Engineering and Applications*, 2021, 57(11): 59.
- [32] 李媛媛, 常庆瑞, 刘秀英, 等. 基于高光谱和 BP 神经网络的玉米叶片 SPAD 值遥感估算[J]. *农业工程学报*, 2016, 32(16): 140.
- LI Y Y, CHANG Q R, LIU X Y, *et al.* Estimation of maize leaf SPAD value based on hyperspectrum and BP neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(16): 140.
- [33] 铁锦程, 赵战营. 基于粒子群优化算法的模型超参调优方法应用研究[J]. *计算机应用与软件*, 2023, 40(10): 152.
- TIE J C, ZHAO Z Y. Study and application of a model hyperparameters optimization method based on particle swarm optimization [J]. *Computer Applications and Software*, 2023, 40(10): 152.
- [34] 唐子竣, 向友珍, 王辛, 等. 利用相关矩阵法优化光谱指数的冬油菜氮素营养诊断[J]. *农业工程学报*, 2023, 39(17): 97.
- TANG Z J, XIANG Y Z, WANG X, *et al.* Nitrogen nutrition diagnosis of winter oilseed rape using spectral indexes optimized by correlation matrix method [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(17): 97.
- [35] KONG W P, HUANG W J, MA L L, *et al.* Estimating vertical distribution of leaf water content within wheat canopies after head emergence [J]. *Remote Sensing*, 2021, 13(20): 4125.
- [36] 王敬湧, 张明珍, 凌华荣, 等. 干旱胁迫下玉米叶片叶绿素含量与含水量高光谱成像反演方法[J]. *智慧农业(中英文)*, 2023, 5(3): 150.
- WANG J Y, ZHANG M J, LING H R, *et al.* A hyperspectral image-based method for estimating water and chlorophyll contents in maize leaves under drought stress [J]. *Smart Agriculture*, 2023, 5(03): 150.
- [37] 赵成, 韩娜娜, 周青云, 等. 基于冠层高光谱的冬小麦植株含水率估算[J/OL]. *麦类作物学报*. <http://kns.cnki.net/kcms/detail/61.1359.S.20231107.1706.037.html>.
- ZHAO C, HAN N N, ZHOU Q Y, *et al.* Estimation of winter wheat plant water content based on canopy hyperspectral [J/OL]. *Journal of Triticeae Crops*. <http://kns.cnki.net/kcms/detail/61.1359.S.20231107.1706.037.html>.
- [38] ZHENG L, CHEN Q, TAO J P, *et al.* Estimation of aboveground biomass for winter wheat at the later growth stage by combining digital texture and spectral analysis [J]. *Agronomy*, 2023, 13(3): 865.
- [39] 谷晓博, 程智楷, 周智辉, 等. 基于特征降维和机器学习的覆膜冬小麦 LAI 遥感反演[J]. *农业机械学报*, 2023, 54(6): 155.
- GU X B, CHENG Z K, ZHOU Z H, *et al.* Remote sensing inversion of leaf area index of mulched winter wheat based on feature downscaling and machine learning [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 155.