

引用格式: Jin Hang, Jing Xia, Gao Yuan, *et al.* GBRT Model for Detecting the Severity of Wheat Stripe Rust by Remote Sensing[J]. Remote Sensing Technology and Application, 2021, 36(2): 411-419. [金航, 竞霞, 高媛, 等. 遥感探测小麦条锈病严重度的GBRT模型研究[J]. 遥感技术与应用, 2021, 36(2): 411-419.]
doi: 10.11873/j.issn.1004-0323.2021.2.0411

遥感探测小麦条锈病严重度的GBRT模型研究

金航^{1,2}, 竞霞¹, 高媛^{1,2}, 刘良云²

(1. 西安科技大学 测绘科学与技术学院, 陕西 西安 710054;

2. 中国科学院遥感与数字地球研究所 数字地球重点实验室, 北京 100094)

摘要: 为了提高小样本数据模型的稳定性, 构建具有更高精度和鲁棒性的小麦条锈病遥感探测模型。首先基于辐亮度和反射率荧光指数方法提取了冠层日光诱导叶绿素荧光(SIF)数据, 然后结合对小麦条锈病病情严重度敏感的反射率光谱指数并基于改进的分类与回归树(CART)——梯度提升回归树(GBRT)算法, 构建了融合反射率和冠层SIF数据的小麦条锈病遥感探测的GBRT模型, 并将其与CART及多元线性回归(MLR)模型进行比较。结果表明: ①反射率导数荧光指数D705/D722、短波红外谷反射率和反射率比值荧光指数R740/R800是影响遥感探测小麦条锈病严重度精度的主控因素, 其中SIF数据的重要性值高于反射率光谱数据, 冠层SIF能够比反射率光谱更加敏感地反映小麦条锈病害信息; ②GBRT模型病情指数(DI)预测值和实测值间的均方根误差(RMSE)比CART和MLR模型分别减小了15.50%和13.49%, 决定系数(R^2)分别提高了6.16%和11.57%, GBRT模型估测DI值更接近于实测值, 且估测结果波动小, 鲁棒性高; CART模型在小样本数据中易将不同特征的数据集划分为同一特征的子集, 预测结果波动较大; MLR模型的预测结果相对比较稳定, 但其预测精度较低。

关键词: GBRT; 日光诱导叶绿素荧光; 反射率光谱; 小麦条锈病; 病情严重度

中图分类号: TP79; S435.121 **文献标志码:** A **文章编号:** 1004-0323(2021)02-0411-09

1 引言

小麦是我国的主要农作物之一, 条锈病是一种随气流传播严重威胁小麦生产的真菌病害^[1]。一旦暴发, 若不及时测报与防治, 将严重影响小麦的产量和质量, 因此预防和监测条锈病的发生及发展已成为学者们研究的重点与难点^[2]。冠层反射光谱能够较好的探测作物群体与结构信息, 对植被生化组分信息的变化比较敏感^[3], 近年来许多学者利用反射光谱数据对小麦条锈病的遥感探测进行了研究并取得了一系列成果^[4-7]。日光诱导叶绿素荧光(Solar-Induced chlorophyll Fluorescence, SIF)是自然光条件下植被吸收的光合有效辐射在满足光合

作用的条件下以长波(650~800 nm)形式发射的叶绿素荧光信号^[8], 能够敏感反映植物光合作用的能力、胁迫状况与生理状况^[9], 已被广泛用于氮素^[10]、焦枯病^[11]和干旱^[12-13]等胁迫探测。张永江等^[14]研究表明, 基于Fraunhofer线原理提取的冠层方向性SIF信息可以反映田间小麦条锈病的发病状况, 然而小麦受条锈病菌侵染后, 其水分及叶绿素含量、光合速率和光能转换率等一些生理生化指标均会发生变化^[15], 仅利用反射率光谱或SIF数据难以全面客观地映射小麦条锈病害的真实状况, 影响小麦条锈病的遥感探测精度, 而协同反射率数据和SIF数据能够提高小麦条锈病的遥感探测精度^[16-17]。

收稿日期: 2019-10-09; 修订日期: 2021-01-26

基金项目: 国家自然科学基金项目资助(41601467)。

作者简介: 金航(1996—), 男, 陕西西安人, 硕士研究生, 主要从事灾害遥感研究。E-mail: jinhang-web@qq.com

通讯作者: 竞霞(1978—), 女, 河南商丘人, 博士, 副教授, 主要从事农业定量遥感研究。E-mail: jingxia1001@163.com

在利用高光谱特征构建小麦病情严重度反演模型时,由于受到数据获取时外界条件等因素的影响,使基于简单回归分析方法建立的数学统计模型在时间维和空间维的普适性较差^[18],因此也有不少学者将机器学习算法应用于小麦病情严重度的遥感探测中,并取得了较好的研究成果。竞霞等^[5]基于冬小麦冠层高光谱特征参量,利用偏最小二乘算法建立了小麦条锈病冠层光谱探测模型,实现了对小麦条锈病早期探测。胡根生等^[6]用粒子群算法优化模型参数,利用优化后的参数建立了小麦白粉病监测模型,取得了较好的监测效果。金秀等^[7]基于深度神经网络有效提取了小麦赤霉病病症的高光谱特征,其建模和验证准确率远高于传统模型。在众多机器学习算法中,分类与回归树(Classification and Regression Tree, CART)算法不需要事先了解用于建模的特征分布类型,对进入模型的初始变量无要求,易于大范围实施。马慧琴^[19]利用 Landsat 8 影像提取的指数特征构建了多种不同的混合小麦病害,通过 CART 模型得到不同特征变量在多变量模型中的贡献并取得较高精度,但该算法存在易过拟合、不稳定、泛化能力弱、预测精度低等不足^[20]。梯度提升回归树(Gradient Boosting Regression Tree, GBRT)算法在 CART 算法的基础上添加梯度迭代(Gradient Boosting, GB)算法和缩减思想(Shrinkage),解决了 CART 算法在小样本数据构建模型中的过拟合等问题,在作物病害的遥感探测中具有一定的优势。

已有研究表明,反射率光谱数据和 SIF 数据均能提供丰富的植被信息,协同反射率数据和冠层 SIF 能够提高小麦条锈病病情严重的估测精度^[16-17]。因此以融合的 SIF 和反射率光谱数据作为特征参量,利用 GBRT 算法构建小麦条锈病病情严重度的遥感探测模型,并将其与 CART 及 MLR 模型进行比较,以确定小麦条锈病遥感探测的适宜方法和模型,为提高小麦条锈病的遥感探测精度提供一种应用参考。

2 数据与方法

2.1 试验设计

试验区位于河北省廊坊市中国农业科学院试验站(39°30'40"N, 116°36'20"E),试验对象为对条锈病比较敏感的小麦品种“铭贤 169 号”,2018 年 4 月 9 日采用浓度为 0.09 mg/ml 的孢子溶液利用喷雾

法对小麦进行条锈病接种。试验区域的小麦分为健康组和染病组,每个试验组的面积为 220 m²,每个组分为 8 个样方(A1-A8、B1-B8、C1-C8、D1-D8),即健康组和染病组各 16 个样方。

2.2 田间冠层光谱测定

试验于 2018 年 5 月 14 日、5 月 18 日和 5 月 24 日采用 ASD FieldSpec 4(350~2 500 nm)地物光谱仪测定小麦在不同病情严重度下的冠层光谱。测量时间为北京时间 11:00~12:30,测量高度距离地面 1.3 m,探头视场角 25°,每区域重复测定 10 次,每次测量前后均用标准 BaSO₄ 参考板进行校正,对测量的冠层光谱数据由公式(1)计算反射率。

$$R = \frac{L_{target}}{L_{board}} \times R_{board} \quad (1)$$

其中: R 为冠层反射率; L_{target} 为目标辐亮度; L_{board} 为参考板辐亮度; R_{board} 为参考板反射率。

2.3 病情指数调查

冠层病情指数(disease index, DI)调查采用 5 点取样法,在每个样方内选取对称的 5 点,每点约 1 m²,随机选取 30 株小麦,分别调查其发病情况。病情严重度参照国家标准“小麦条锈病测报技术规范”(GB/T 15795-2011)进行量化,单叶严重度分 9 个等级,即:0、1%、10%、20%、30%、45%、60%、80% 和 100%,对应的级值代表值为 1~9,利用公式(2)计算小麦群体病情指数。

$$DI = \frac{\sum(x \times f)}{n \times \sum f} \times 100\% \quad (2)$$

其中:DI 为病情指数; x 为各梯度级值; f 为各梯度的叶片数; n 为梯度值。

2.4 特征参量提取

为了利用 GBRT 算法构建小麦条锈病病情严重度估测模型,参考已有的研究成果^[21],分别利用反射率光谱与冠层 SIF 数据提取了 24 个反射率光谱指数及叶绿素荧光特征参数^[14, 22-29](表 1)。

由于光谱采样间隔的离散性,采用差分方法近似计算了一阶微分光谱:

$$\rho'(\lambda_i) = [\rho(\lambda_{i+1}) - \rho(\lambda_{i-1})] / 2\Delta\lambda \quad (3)$$

其中: λ_i 为各波段波长; $\rho'(\lambda_i)$ 为一阶微分光谱; $\Delta\lambda$ 为波长 λ_{i+1} 到 λ_{i-1} 的间隔。

在提取叶绿素荧光特征参量时,分别利用辐亮度和反射率荧光指数 2 种方法^[14]计算了小麦条锈病不同病情严重度下的冠层 SIF 强度。基于辐亮度的冠层 SIF 提取算法是根据夫琅和费暗线原理,通过计算自然光照条件下太阳光激发的荧光对夫琅和

表 1 高光谱及荧光特征参数与定义

Table 1 Hyperspectral and fluorescence characteristic parameters and definitions

光谱特征变量	参数说明
蓝边面积 ^[24] Blue edge area(X_1)	蓝光范围(490~530 nm)内的反射率光谱对波长的积分值
蓝边斜率 Blue edge slope(X_2)	蓝边位置对应的一阶微分值
黄边面积 ^[24] Yellow edge area (X_3)	黄光范围(560~590 nm)内的反射率光谱对波长的积分值
黄边斜率 Yellow edge area(X_4)	黄边位置对应的一阶微分值
红边面积 ^[24] Red edge area (X_5)	红光范围(680~760 nm)内的反射率光谱对波长的积分值
红边斜率 ^[26] Red edge slope(X_6)	红边位置对应的一阶微分值
红边位置 ^[25] Red edge position (X_7)	红光范围内一阶微分光谱对应的一阶微分值
绿峰位置 ^[24] Green peak position (X_8)	波长在 510~560 nm 波段反射率最大值所对应的波长
红谷反射率 ^[27] Red Valley reflectivity (X_9)	波长在 640~700 nm 波段反射率最小值
红谷面积 Red Valley area(X_{10})	波长在 640~700 nm 内原始光谱曲线所包围的面积
短波红外谷反射率 ^[28] Short-Wave Infrared Valley Reflectivity (X_{11})	波长在 1400~1800 nm 波段反射率最小值
红光区反射率最大值 ^[29] Red Light Area Reflectivity Maximum (X_{12})	红光区(620~680 nm)反射率最大值
红边反射率最大值 Red Edge Reflectance Maximum (X_{13})	红边(680~780 nm)反射率最大值
近红外区反射率最大值 Near-Infrared Reflectance Maximum (X_{14})	近红外区(780~1100 nm)反射率最大值
黄光区反射率总和 Yellow Light Area Reflectance Sum (X_{15})	黄光区(560~590 nm)反射率总和
红光区反射率总和 Red Light Area Reflectance Sum (X_{16})	红光区(620~680 nm)反射率总和
红边反射率总和 Red Edge Reflectance Sum (X_{17})	红边(680~780 nm)反射率总和
近红外区反射率总和 Near-Infrared Reflectance Sum (X_{18})	近红外区(780~1100 nm)反射率总和
绿峰反射高度 ^[29] Green Peak Reflection Height (X_{19})	$1 - \frac{R_{490} + \frac{R_{670} - R_{490}}{670 - 490} (560 - 490)}{R_{560}}$
红谷吸收深度 ^[29] Red Valley Absorption Depth (X_{20})	$1 - \frac{R_{670}}{R_{560} + \frac{R_{760} - R_{560}}{760 - 560} (670 - 560)}$
短波红外谷吸收深度 ^[29] Short-Wave Infrared Valley Absorption Depth (X_{21})	$1 - \frac{R_{1450}}{R_{1400} + \frac{R_{1670} - R_{1400}}{1670 - 1400} (1450 - 1400)}$
反射率荧光比值指数 ^[22] Reflectance Fluorescence Ratio Index (X_{22})	R740/R800
反射率荧光导数指数 ^[23] Reflectance Fluorescence Derivative Index (X_{23})	D705/D722
日光诱导叶绿素荧光 ^[14] Solar-Induced Chlorophyll Fluorescence (X_{24})	$\overline{F_{in}} = \frac{L_{in} \times (\omega_{left} \times I_{left} + \omega_{right} \times I_{right}) - I_{in} \times (\omega_{left} \times L_{left} + \omega_{right} \times L_{right})}{(\omega_{left} \times I_{left} + \omega_{right} \times I_{right}) - I_{in}}$

费“井”的填充程度估算冠层 SIF 强度。关于填充程度的计算,目前已有多种算法,在此采用 Maier 等^[30]

提出的鲁棒性较好的 3FLD 算法估测 O2-A(760 nm)波段下冠层 SIF 强度。

$$\overline{F_{in}} = \frac{L_{in} \times (\omega_{left} \times I_{left} + \omega_{right} \times I_{right}) - I_{in} \times (\omega_{left} \times L_{left} + \omega_{right} \times L_{right})}{(\omega_{left} \times I_{left} + \omega_{right} \times I_{right}) - I_{in}} \quad (4)$$

其中: I_{in} 为夫琅和费线内的太阳辐照度光谱强度; L_{in} 为夫琅和费线内的植被冠层反射的辐亮度光谱强度; ω_{left} 、 ω_{right} 表示吸收线左右两参考波段所占权重; I_{left} 、 I_{right} 表示吸收线左右的太阳辐照度光谱强度; L_{left} 、 L_{right} 表示吸收线左右的植被冠层反射辐亮度光谱强度。

利用反射率荧光指数计算冠层 SIF 强度是通过分析荧光对 650~800 nm 红边区域反射率的影响构建的荧光光谱指数,主要包括反射率比值荧光指数和反射率导数荧光指数等^[14],实验主要采用了反射率比值荧光指数 R740/R800^[23]和反射率导数荧光

指数 D705/D722^[30]。

2.5 研究方法

首先按照小麦条锈病病情严重度进行分层,每层依据 3:1 的比例从 48 个样本点中随机选取 36 个样本作为训练数据进行模型构建,剩余 12 个样本用于检验数据评价模型精度。然后采用格网搜索方法确定 GBRT 模型的弱学习器数量(n_trees),学习速率(learning_rate)以及决策树深度(max_depth)3 个模型参数的最优值。最后以不同 DI 值下的反射率光谱特征及冠层 SIF 为自变量,以小麦条锈病病情严重度为响应变量,基于 Python 平台构建了融合

反射率和冠层 SIF 数据的小麦条锈病遥感探测的 GBRT 模型。

2.5.1 分类与回归树(CART)

CART 是一种基于决策树的机器学习方法,该方法通过二分递归的方式自动选择自变量,将原始数据集划分为具有同质特点的多个子集,利用决策树表达变量间的复杂关系^[31]。CART 模型的构建主要包括树的生成和树的剪枝两部分^[32]。树的生成主要过程为对训练数据集二元分割为两个子集,通过对子集再分割不断递归生成树,使生成的决策树尽量大;树的剪枝主要包括预剪枝和后剪枝两种方法,其中后剪枝方法修剪过程比预剪枝方法更为直观形象,而且该方法同时考虑了决策树的预测误差以及树的复杂程度,使模型具有更好的泛化能力^[33]。实验采用后剪枝进行决策树的修剪,剪枝规则选用最小代价复杂度规则,其计算如公式(5)所示:

$$\alpha = \frac{R(t) - R(T)}{N(T) - 1} \quad (5)$$

其中: $R(t)$ 为叶子节点的误差代价; $R(T)$ 为子树的误差代价; $N(T)$ 为子树节点个数。

2.5.2 梯度提升回归树(GBRT)

GBRT 算法能够弥补 CART 算法对小样本数据易过拟合、不稳定、预测精度低的不足。该算法是由 Friedman^[34]提出的一种迭代的决策树算法,它由 CART、GB 算法以及缩减思想 3 部分组成^[35]。该算法的基本思想是利用 Boosting 方法将多个预测精度低的弱学习器多次迭代构成一个预测精度高的强学习器,即通过再一次学习减少上一次模型的残差,使得下一次生成的模型有更小的误差,梯度迭代使组合模型不断得到改进,是一种决策树集成学习算法;而缩减的本质是为了降低模型学习速率,防止过拟合,即它不完全信任每一棵残差树,采用逐渐逼近的方式通过多棵树残差进行学习^[36]。

对于连续的数据类型,损失函数采用 Boosting 中的经典损失函数,即误差平方和,其计算方法如式(6)所示,经过 M 次迭代后的 GBRT 模型预测结果如式(7)所示:

$$L(y, f(x)) = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (6)$$

$$f(x) = \sum_{i=0}^M f_i(x) \quad (7)$$

其中: x 为 y 分别为输入变量和输出变量; i 为迭代次数; $f_0(x)$ 为初始预测值, $f(x)$ 为模型预测值。

提高 GBRT 模型预测精度的关键是残差的计算,本文利用 Friedman^[34]提出的基于损失函数负梯度残差的计算方法,即利用损失函数的负梯度作为提升树算法中的残差的近似值。第 m 轮的第 i 个样本的损失函数负梯度 g_{im} 计算公式如式(8)所示:

$$g_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x_i) = f_{m-1}(x_i)} \quad (8)$$

$f_m(x)$ 可通过 β_m 和 h_m 计算得到,其中 β_m 表示每次最优迭代步长,计算方法如公式(9)所示,而 $h_m(x)$ 为第 m 次迭代所创建的决策树。通过 m 次弱学习器迭代后的预测结果如式(10)所示:

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^n L(y_i, f_{m-1}(x) + \beta h_m(x)) \quad (9)$$

$$f_m(x) = f_{m-1}(x) + \beta_m h_m(x) \quad (10)$$

2.5.3 多元线性回归(MLR)

多元线性回归(Multiple Linear Regression, MLR)是根据多个输入参数求出相应回归系数,将回归系数与对应自变量乘积后累加得到因变量,利用线性模型建立因变量与多个自变量之间关系的一种建模方法。其中,自变量须与因变量有密切线性相关关系并具有一定的互斥性。多元线性回归模型如式(11)所示:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \mu \quad (11)$$

其中: Y 为被解释变量, $X_j (j=1, 2, \cdots, n)$ 为解释变量, $\beta_j (j=0, 1, 2, \cdots, k)$ 为回归系数, μ 为随机误差且遵从标准正态分布。

3 结果与分析

3.1 模型构建

3.1.1 CART 模型

为防止 CART 回归树的过拟合,通过计算模型复杂度参数、相对误差及其交叉验证相对误差随决策树分割节点数量增加的变化情况,如表 2 所示。

由此得到 10 折交叉验证下的 X-Val 相对误差和模型复杂度参数 CP 的关系,以模型复杂性小、交叉验证误差低的标准作为 CART 模型最优修剪准则。依据上述准则确定了本文中 CART 模型的最优修剪规则为 $X_{\text{error}}=0.349$, $CP=0.01$ 。

3.1.2 GBRT 模型

在利用 Python 平台建立 GBRT 模型的过程中,需要确定弱学习器数量,学习速率以及决策树深度 3 个模型参数,其中弱学习器数量应当在误差趋于稳定后越小越好,决策树深度为避免过拟合现象不

表 2 CART 模型复杂度参数(CP)、相对误差(REL_{error})及其交叉验证相对误差(X_{error})

Table 2 CP, REL _{error} and X _{error} of CART			
nsplit	CP	REL _{error}	X _{error}
0	0.678 919	1	1.063 38
1	0.143 475	0.321 080 6	0.507 58
2	0.070 193	0.177 605 9	0.460 05
3	0.043 176	0.107 412 9	0.469 07
4	0.016 467	0.064 237	0.357 14
5	0.010 215	0.031 303 3	0.349 4
6	0.009 33	0.021 088 3	0.359 14
7	0.003 453	0.011 758	0.402 98
8	0.002 919	0.008 304 9	0.389 39

应过大,学习速率通常取 $v\leq 0.1$ 。实验采用格网搜索方法确定遥感探测小麦条锈病 GBRT 模型的最优参数。通过格网搜索得到小麦条锈病预测 DI 值与实测 DI 值的平均绝对误差随分类回归树数量以及深度的变化趋势如图 1 和图 2 所示,由图可知,当 GBRT 模型中弱学习器的数量达到 200 后,其预测结果的平均绝对误差低且趋于稳定;当树的深度为 4 时,其预测结果的平均误差最小;通过对训练样本的多次仿真发现将学习速率设置为默认值 0.1 时模型的精度较高,在本模型中, {n_trees=200, learning_rate=0.1, max_depth=4} 参数组合为最优组合。

3.1.3 MLR 模型

在构建小麦条锈病严重度遥感探测的多元线性回归模型时,自变量的优选依据为特征参量和 DI 之间的显著性水平与方差膨胀因子 (Variance Inflation Factor, VIF),其中显著性水平设定为 $p\leq 0.05$,而方差膨胀因子的阈值设定为 10,这是因为 VIF 越大,变量之间共线性越严重。当 $0<VIF<10$,不存在多重共线性;当 $10\leq VIF<100$,存在较强的多重共线性;当 $VIF\geq 100$,存在严重多重共线性^[37]。实验将 VIF 值大于等于 10 的自变量舍去,筛选得到两种不存在多重共线性的变量 X_7 、 X_{24} ,以此为自变量建立 MLR 模型。

GBRT、CART 和 MLR 模型对小麦条锈病严重度的预测结果如图 3 所示,图中 Test 为测试数据,GBRT 为梯度提升回归树预测结果,CART 为分类与回归树预测结果,MLR 为多元线性回归预测结果。由图 3 可以看出,GBRT 模型估测小麦条锈病病情严重度总体上效果良好,更加接近实测值,估测结果波动小,鲁棒性高;CART 模型预测结果处于较大波动状态,这是由于 CART 模型在小样本数

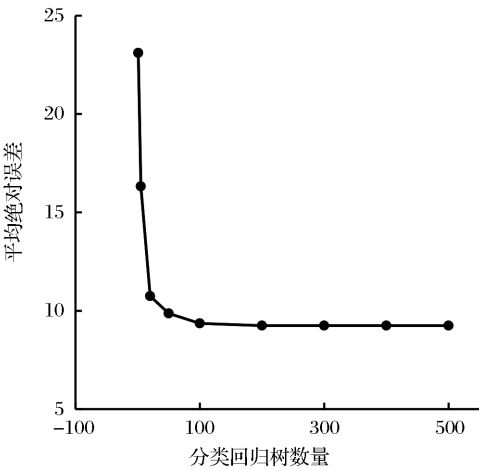


图 1 平均绝对误差随分类回归树数量的变化趋势
Fig.1 Trend of MAE varying with number of CART

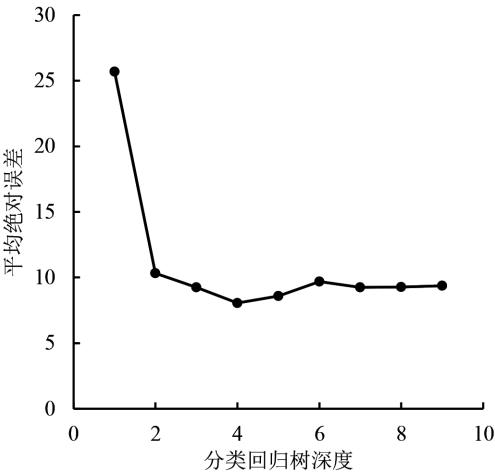


图 2 平均绝对误差随分类回归树深度的变化趋势
Fig.2 Trend of MAE varying with depth of CART

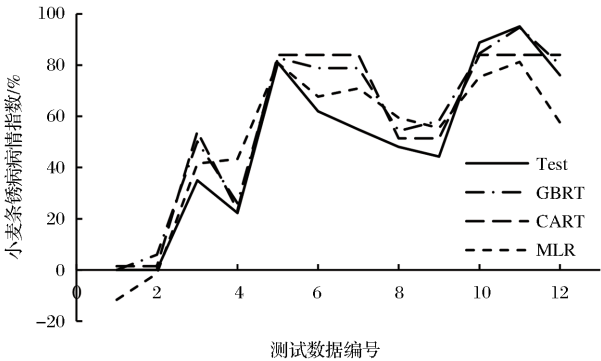


图 3 不同预测模型对条锈病病情指数预测结果
Fig.3 Prediction results of stripe rust disease index by different prediction models

据中易将不同特征的数据集划分为同一特征的子集,从而影响估测的准确度;MLR 所建估测模型相对比较稳定,但其预测精度较低。综上所述,GBRT 模型对病情严重度具有良好的预测能力、泛化能力

和鲁棒性,在小麦条锈病的遥感估测中具有更好的适用性。

3.2 模型精度评价

为了评价小麦条锈病病情严重度估测模型的可靠性,采用模型估测DI值与实测DI间的决定系数(determination coefficient, R^2)和均方根误差(Root Mean Square Error, RMSE)^[16]2个指标进行模型精度评价,其中 R^2 和RMSE计算如公式(12)和(13)所示。

$$R^2 = \frac{(\sum (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i))^2}{\sum (y_i - \bar{y}_i)^2 \sum (\hat{y}_i - \bar{\hat{y}}_i)^2} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (13)$$

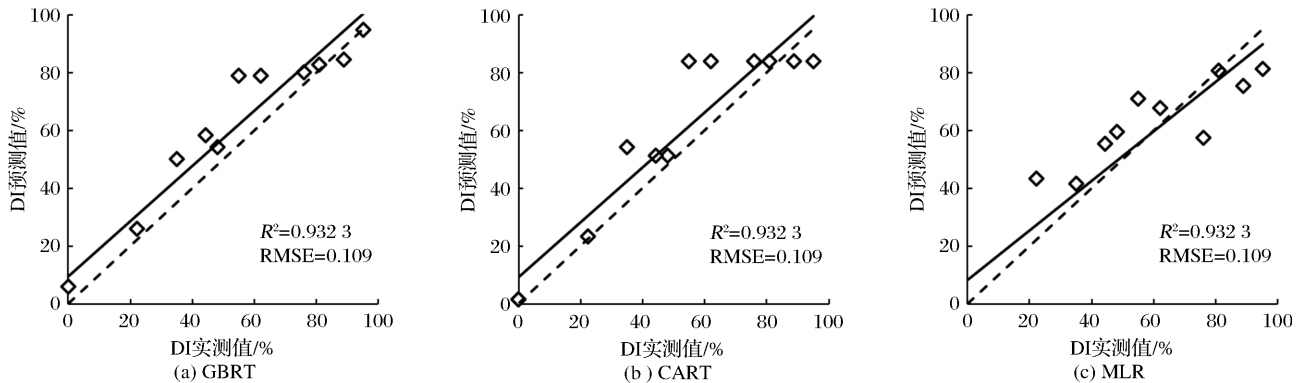


图4 不同模型的预测DI值与实测DI值拟合

Fig.4 Comparison of the estimated DI and real DI using different model

3.3 小麦病情严重度主控因素分析

GBRT模型能够通过计算特征的重要程度,发现特征间的高阶关系。特征重要性通过不同特征在单棵树中重要程度的平均值进行衡量,如公式(14)所示。

$$\hat{J}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{J}_j^2(T_m) \quad (14)$$

其中: M 表示树的数量,特征 j 在单棵树中的重要度通过该特征分裂后损失的减少值进行计算。

通过Python平台计算特征重要性,得到影响小麦病情严重度的主控因素为反射率荧光导数指数D705/D722、短波红外谷反射率和反射率荧光比值指数R740/R800。

荧光强度能够敏感反映病害引起的生理变化及其受胁迫状态^[38],起到植物健康状态“探针”作用^[39];短波红外谷处于水分敏感波段处,随着小麦条锈病病情严重度的增加,小麦水分严重损失,使其短波红外谷反射率变化差异非常显著,这与已有

其中: n 为检验样本个数; y_i 和 \bar{y}_i 分别为病情严重度实测值与实测值的平均值; \hat{y}_i 和 $\bar{\hat{y}}_i$ 分别为模型预测值与预测值的平均值。

利用相同的验证样本分别利用公式(12)和(13)计算3种不同模型预测DI值与实测DI值间的 R^2 和RMSE,结果如图4所示,图中实线为不同模型预测DI值和实测DI值之间的拟合线,虚线为1:1关系线。由图4可以看出GBRT模型预测DI值和实测DI值间的 R^2 比CART模型提高了6.16%,比MLR模型提高了11.57%,而RMSE则比CART模型减小了15.50%,比MLR模型减小了13.49%。GBRT模型估测精度优于CART模型和MLR回归模型,更适合于小麦条锈病的遥感探测。

研究结果相一致^[40-41]。

4 结 语

论文利用改进CART算法——GBRT算法,以格网搜索的方式得到模型参数的最优组合,构建了遥感估测小麦条锈病病情严重度模型,并将之与CART、MLR模型进行比较,从而确定高光谱遥感估测小麦条锈病严重度的适宜算法和模型。

(1)GBRT作为以Boosting为策略的集成学习,在保证偏差的同时选择深度较浅的决策树便能达到较高的精度,通过格网搜索得到学习速率为0.1,弱学习器数量为200,树的深度为4时的参数组合为GBRT模型最优组合,可作为GBRT算法估测小麦条锈病严重度的参数设置依据。

(2)通过计算GBRT特征的重要程度,得到重要性前3的特征参数分别为反射率荧光导数指数D705/D722、短波红外谷反射率和反射率荧光比值指数R740/R800,说明冠层SIF比反射率光谱能更

加敏感地反映小麦条锈病害信息。这是因为虽然反射率光谱能够敏感反映作物色素含量变化^[42-43],但受土壤等非绿色景观成分背景噪声的影响较大^[44],且对作物光合活性不敏感,而冠层SIF则受背景噪声的影响较小^[45],能够灵敏反映作物光合生理上的变化,SIF和光合作用之间的直接联系可提高植物胁迫遥感监测精度^[44]。

(3)GBRT模型对小麦条锈病严重度的估测精度更接近于实测值,估测结果波动小,鲁棒性高;由于CART模型在小样本数据中易将不同特征的数据集划分为同一特征的子集,影响了病情严重度的估测精度,导致CART模型预测结果处于较大波动状态;MLR算法所建模型的估测结果相对比较稳定,但其预测精度较低。GBRT模型对病情严重度具有更好的预测能力、泛化能力和鲁棒性,更适用于小麦条锈病的遥感探测。

(4)实验在利用遥感数据构建小麦条锈病严重度估测模型时,并未考虑小麦生育期及其品种对冠层光谱数据的影响,也未考虑反射率光谱数据的目标光谱和背景光谱的混合问题^[46],因此该模型的外延性还需进一步研究。

(5)反射率和荧光的变化是小麦条锈病发生的一种表象特征,不同的胁迫均能引起反射率和荧光数据的变化,仅仅利用反射率和荧光数据只能够实现小麦条锈病严重度探测,如何依据相关的植保和农学知识,综合利用病害发生的时相信息和温湿度信息以及反射率和叶绿素荧光数据实现小麦条锈病遥感探测的逆向识别还有待研究。

参考文献(References):

- [1] Yang Yuheng, Su Qiaoyan, Wang Ze, *et al.* Occurrence Regularity and Influence Factors of Wheat Strip Rust in Chongqing [J]. Journal of Northwest A&F University (Natural Science Edition), 2016, 44(9):151-157.[杨宇衡, 宿巧燕, 王泽, 等. 重庆市小麦条锈病发生规律和影响因素分析[J]. 西北农林科技大学学报(自然科学版), 2016, 44(9):151-157.]
- [2] Huang Muiyi, Wang Jihua, Huang Wenjiang, *et al.* Hyperspectral Character of Stripe Rust on Winter Wheat and Monitoring by Remote Sensing [J]. Transactions of the CSAE, 2003, 19(6):154-158.[黄木易, 王纪华, 黄文江, 等. 冬小麦条锈病的光谱特征及遥感监测[J]. 农业工程学报, 2003, 19(6):154-158.]
- [3] Davoud A, Mohammad M, Alfredo H. Developing Two Spectral Disease Indices for Detection of Wheat Leaf Rust (Pucciniatriticina) [J]. Remote Sensing, 2014, 6(6):4723-4740.
- [4] Gao Yuan, Jing Xia, Liu Liangyun, *et al.* Remote Sensing Monitoring of Wheat Stripe Rust based on Multiple Kernel SVM [J]. Journal of Triticeae Crops, 2020(1):1-9.[高媛, 竞霞, 刘良云, 等. 基于多核支持向量机的小麦条锈病遥感监测研究[J]. 麦类作物学报, 2020(1):1-9.]
- [5] Jing Xia, Lü Xiaoyan, Zhang Chao, *et al.* Early Detection of Winter Wheat Stripe Rust based on SIF-PLS Model [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6):191-197.[竞霞, 吕小艳, 张超, 等. 基于SIF-PLS模型的冬小麦条锈病早期光谱探测[J]. 农业机械学报, 2020, 51(6):191-197.]
- [6] Hu Gensheng, Wu Wentian, Huang Wenjiang, *et al.* Application of PSO-LSSVM in Wheat Powdery Mildew Monitoring [J]. Remote Sensing Technology and Application, 2017, 32(2):299-304.[胡根生, 吴问天, 黄文江, 等. 粒子群优化的最小二乘支持向量机在小麦白粉病监测中的应用[J]. 遥感技术与应用, 2017, 32(2):299-304.]
- [7] Jin Xiu, Lu Jie, Fu Yunzhi, *et al.* A Classification Method for Hyperspectral Imaging of Fusarium Head Blight Disease Symptom based on Deep Convolutional Neural Network [J]. Acta Agriculturae Zhejiangensis, 2019, 31(2):315-325.[金秀, 卢杰, 傅运之, 等. 基于深度卷积神经网络的小麦赤霉病高光谱病症点分类方法[J]. 浙江农业学报, 2019, 31(2):315-325.]
- [8] Hu Jiaochan, Liu Liangyun, Liu Xinjie. Assessing Uncertainties of Sun-induced Chlorophyll Fluorescence Retrieval Using FluorMOD Model [J]. Journal of Remote Sensing, 2015, 19(4):594-608.[胡姣婵, 刘良云, 刘新杰. FluorMOD模拟叶绿素荧光夫琅和费暗线反演算法不确定性分析[J]. 遥感学报, 2015, 19(4):594-608.]
- [9] Liu Liangyun. Principle and Application of Vegetation Quantitative Remote Sensing [M]. Beijing: Science Press, 2014.[刘良云. 植被定量遥感原理与应用[M]. 北京: 科学出版社, 2014.]
- [10] Murchie E H, Lawson T. Chlorophyll Fluorescence Analysis: A Guide to Good Practice and Understanding Some New Applications [J]. Journal of Experimental Botany, 2013, 64(13):3983-3998.
- [11] Zarco-Tejada P J, Camino C, Beck P S A, *et al.* Previsual Symptoms of Xylella Fastidiosa Infection Revealed in Spectral Plant-trait Alterations [J]. Nature Plants, 2018, 4(7):432-439.
- [12] Sun Y, Fu R, Dickinson R, *et al.* Drought Onset Mechanisms Revealed by Satellite Solar-induced Chlorophyll Fluorescence: Insights from Two Contrasting Extreme Events [J]. Journal of Geophysical Research: Biogeosciences, 2015, 120(11):2427-2440.
- [13] Liu L, Yang X, Zhou H, *et al.* Evaluating the Utility of Solar-induced Chlorophyll Fluorescence for Drought Monitoring by Comparison with NDVI Derived from Wheat Canopy [J]. Science of the Total Environment, 2018, 625:1208-1217.
- [14] Zhang Yongjiang, Huang Wenjiang, Wangjihua, *et al.* Chlorophyll Fluorescence Sensing to Detect Stripe Rust in Wheat (Triticum Aestivum L.) Fields based on Fraunhofer Lines [J]. Scientia Agricultura Sinica, 2007(1):78-83.[张永江, 黄文江, 王纪华, 等. 基于Fraunhofer线的小麦条锈病荧光遥感探测[J]. 中国农业科学, 2007(1):78-83.]

- [15] Liu Qi, Wang Cuicui, Wang Rui, *et al.* Hyperspectral Qualitative Identification on Latent Period of Wheat Stripe Rust[J]. Journal of Plant Protection, 2018, 45(1): 153-160.[刘琦, 王翠翠, 王睿, 等. 潜育期小麦条锈菌的高光谱定性识别[J]. 植物保护学报, 2018, 45(1): 153-160.]
- [16] Jing Xia, Bai Zongfan, Gao Yuan, *et al.* Wheat Stripe Rust Monitoring by Random Forest Algorithm Combined with SIF and Reflectance Spectrum[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(13): 154-161.[竞霞, 白宗璠, 高媛, 等. 利用随机森林法协同SIF和反射率光谱监测小麦条锈病[J]. 农业工程学报, 2019, 35(13): 154-161.]
- [17] Chen Siyuan, Jing Xia, Dong Yingying, *et al.* Detection of Wheat Stripe Rust Using Solar-induced Chlorophyll Fluorescence and Reflectance Spectral Indices[J]. Remote Sensing Technology and Application, 2019, 34(3): 511-520.[陈思媛, 竞霞, 董莹莹, 等. 基于日光诱导叶绿素荧光与反射率光谱的小麦条锈病探测研究[J]. 遥感技术与应用, 2019, 34(3): 511-520.]
- [18] Wang Jing, Jing Yuanshu, Huang Wenjiang, *et al.* Comparative Research on Estimating the Severity of Yellow Rust in Winter Wheat[J]. Spectroscopy and Spectral Analysis, 2015, 35(6): 1649-1653.[王静, 景元书, 黄文江, 等. 冬小麦条锈病严重度不同估算方法对比研究[J]. 光谱学与光谱分析, 2015, 35(6): 1649-1653.]
- [19] Ma Huiqin. Remote Sensing Monitoring and Forecasting Models of Wheat Powdery Mildew based on Multi-Source Data [D]. Nanjing: Nanjing University of Information Science and Technology, 2017.[马慧琴. 基于多源数据的小麦白粉病遥感监测与预测模型研究[D]. 南京: 南京信息工程大学, 2017.]
- [20] Liu Yunxiang, Wu Hao. Establishment of Water Bloom Warning Model based on Improved CART Decision Tree[J]. China Rural Water and Hydropower, 2018(1): 26-28.[刘云翔, 吴浩. 基于改进CART决策树建立水华预警模型[J]. 中国农村水利水电, 2018(1): 26-28.]
- [21] Huang Wenjiang, Zhang Jingcheng, Shi Yue, *et al.* Progress in Monitoring and Forecasting of Crop Pests and Diseases by Remote Sensing[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2018, 10(1): 30-43.[黄文江, 张竞成, 师越, 等. 作物病虫害遥感监测与预测研究进展[J]. 南京信息工程大学学报(自然科学版), 2018, 10(1): 30-43.]
- [22] Zarco-Tejada P J, Miller J R, Mohammed G H, *et al.* Chlorophyll Fluorescence Effects on Vegetation Apparent Reflectance: I. Leaf-level Measurements and Model Simulation[J]. Remote Sensing of Environment, 2000, 74: 582-595.
- [23] Zarco-Tejada P J, Pushnik J C, Dobrowski S Z, *et al.* Steady-state Chlorophyll a Fluorescence Detection from Canopy Derivative Reflectance and Double-peak Red-edge Effects [J]. Remote sensing of Environment, 2003, 84(2): 283-294.
- [24] Bi Yinli, Sun Jiangtao, Ypyszhn, *et al.* Hyperspectral Characterization and Nutrition Condition of Maize Inoculated with Arbuscular Mycorrhiza in Different Phosphorus Levels [J]. Journal of China Coal Society, 2016, 41(5): 1227-1235.[毕银丽, 孙江涛, Ypyszhn, 等. 不同施磷水平下接种菌根玉米营养状况及光谱特征分析[J]. 煤炭学报, 2016, 41(5): 1227-1235.]
- [25] Zhang Yonghe, Chen Wenhui, Guo Qiaoying, *et al.* Hyperspectral Estimation Models for Photosynthetic Pigment Contents in Leaves of Eucalyptus [J]. Acta Ecologica Sinica, 2013, 33(3): 876-887.[张永贺, 陈文惠, 郭乔影, 等. 桉树叶片光合色素含量高光谱估算模型[J]. 生态学报, 2013, 33(3): 876-887.]
- [26] Yang X H, Wang F M, Huang J F, *et al.* Comparison between Radial Basis Function Neural Network and Regression Model for Estimation of Rice Biophysical Parameters Using Remote Sensing[J]. Pedosphere, 2009, 19(2): 176-188.
- [27] Huang J F, Blackburn G A. Optimizing Predictive Models for Leaf Chlorophyll Concentration based on Continuous Wavelet Analysis of Hyperspectral Data[J]. International journal of remote sensing, 2011, 32(24): 9375-9396.
- [28] Huang Xiaojun, Xie Yaowen, Bao Yuhai. Spectral Detection of Damaged Lever of Larch Affected by Jas's Larch Inchworm [J]. Spectroscopy and Spectral Analysis, 2018(3): 905-911.[黄晓君, 颀耀文, 包玉海. 落叶松受雅氏落叶松尺蠖危害程度光谱检测[J]. 光谱学与光谱分析, 2018(3): 905-911.]
- [29] Zhang Sulan, Tan Ju, Tang Xiaodong, *et al.* Spectral Characteristics and Evaluation Model of Pinus Massoniana Suffering from Bursaphelenchus Xylophilus Disease [J]. Spectroscopy and Spectral Analysis, 2019, 39(3): 865-872.[张素兰, 覃菊, 唐晓东, 等. 松材线虫危害下马尾松光谱特征与估测模型研究[J]. 光谱学与光谱分析, 2019, 39(3): 865-872.]
- [30] Maier S W, Günther K P, Stellmes M. Sun-induced Fluorescence: A New Tools for Precision Farming[C]// Digital Imaging and Spectral Techniques: Applications to Precision Agriculture and Crop Physiology. Madison: American Society of Agronomy, 2003: 209-222.
- [31] Yu Wenqi, Dai Xue, Yang Ying, *et al.* The Relationship between Water-level Fluctuation Factors and the Distribution of Carex in Floodplain Grassland Around Poyang Lake[J]. Journal of Lake Sciences, 2018, 30(6): 204-212.[于文琪, 戴雪, 杨颖, 等. 基于CART模型的鄱阳湖草滩苔草分布与水位波动要素关系[J]. 湖泊科学, 2018, 30(6): 204-212.]
- [32] Han Jiaqi, Mao Kebiao, Ge Feifan, *et al.* Application of Classification and Regression Tree in Soil Moisture Estimation[J]. Remote Sensing Information, 2018, 33(3): 49-56.[韩家琪, 毛克彪, 葛非凡, 等. 分类回归树算法在土壤水分估算中的应用[J]. 遥感信息, 2018, 33(3): 49-56.]
- [33] Huang Xiaojun, Xie Yaowen, Wei Jiaojiao, *et al.* Automatic Recognition of Desertification Information based on the Pattern of Change Detection-CART Decision Tree[J]. Journal of Catastrophology, 2017, 32(1): 36-42.[黄晓君, 颀耀文, 卫娇娇, 等. 基于变化检测-CART决策树模式自动识别沙漠化信息[J]. 灾害学, 2017, 32(1): 36-42.]
- [34] Friedman J H. Greedy Function Approximation: A Gradient Boosting Machine[J]. The Annals of Statistics, 2001, 29(5): 1189-1232.
- [35] Qiu Xin, Hong Haoyu, Yang Qing, *et al.* Prediction of Temperature of Asphalt Pavement Surface based on APRIORI-GBDT Algorithm.[J]. Journal of Highway and Transportation Research and Development, 2019, 36(5): 1-10, 19.[邱欣, 洪皓珩, 杨青, 等. 基于APRIORI-GBDT算法的沥青路面路表温度预测[J]. 公路交通科技, 2019, 36(5): 1-10, 19.]

- [36] Zhang Jialong, Xu Hui, Lu Chi. Estimating Above Ground Biomass of *Pinus Densata* based on Landsat8 OLI and Gradient Boost Regression Tree[J]. Journal of Northeast Forestry University, 2018, 46(8):27-32.[张加龙, 胥辉, 陆驰. 应用 Landsat8 OLI 和 GBRT 对高山松地上生物量的估测[J]. 东北林业大学学报, 2018, 46(8):27-32.]
- [37] Zuur A F, Ieno E N, Elphick C S. A Protocol for Data Exploration to Avoid Common Statistical Problems[J]. Methods in Ecology and Evolution, 2010, 1(1):3-14.
- [38] Subhash N, Subhash N, Ravi V, *et al.* Detection of Mosaic Virus Disease in Cassava Plants by Sunlight-induced Fluorescence Imaging: A Pilot Study for Proximal Sensing[J]. International Journal of Remote Sensing, 2015, 36(11):2880-289.
- [39] Krause G H, Weis E. Chlorophyll Fluorescence and Photosynthesis: The Basics [J]. Annual Review of Plant Biology, 1991, 42(1): 313-349.
- [40] Liu L, Zhao J, Guan L, *et al.* Tracking Photosynthetic Injury of Paraquat-treated Crop Using Chlorophyll Fluorescence from Hyperspectral Data [J]. European Journal of Remote Sensing, 2013, 46(1):459-473.
- [41] Huang Mui, Huang Wenjiang, Liu Liangyun, *et al.* Spectral Reflectance Feature of Winter Wheat Single Leaf Infected with Stripe Rust and Severity Level Inversion [J]. Transactions of the CSAE, 2004, 20(1):176-180.[黄木易, 黄文江, 刘良云, 等. 冬小麦条锈病单叶光谱特性及严重度反演[J]. 农业工程学报, 2004, 20(1):176-180.]
- [42] Beck P S A, Goetz S J. Satellite Observations of High Northern Latitude Vegetation Productivity Changes between 1982 and 2008: Ecological Variability and Regional Differences[J]. Environmental Research Letters, 2011, 6(4): 045501. doi: iop.science.iop.org/1748-9326/6141045501.
- [43] Gamon J A, Kovalchuck O, Wong C Y S, *et al.* Monitoring Seasonal and Diurnal Changes in Photosynthetic Pigments with Automated PRI and NDVI Sensors[J]. Biogeosciences, 2015, 12(13):4149-4159.
- [44] Song L, Guanter L, Guan K, *et al.* Satellite Sun-induced Chlorophyll Fluorescence Detects Early Response of Winter Wheat to Heat Stress in the Indian Indo-gangetic Plains[J]. Global Change Biology, 2018-05-10. doi:10.1111/gcb.14302.
- [45] Baker, Neil R. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo[J]. Annual Review of Plant Biology, 2008, 59(1): 89-113.
- [46] Bai Zongfan, Jing Xia, Zhang Teng, *et al.* Canopy SIF Synergize with Total Spectral Reflectance Optimized by the MDBP-SO Algorithm to Monitor Wheat Stripe Rust[J]. Acta Agronomica Sinica, 2020, 46(8): 1248-1257.[白宗璠, 竞霞, 张腾, 等. MDBPSO 算法优化的全波段光谱数据协同冠层 SIF 监测小麦条锈病[J]. 作物学报, 2020, 46(8):1248-1257.]

GBRT Model for Detecting the Severity of Wheat Stripe Rust by Remote Sensing

Jin Hang^{1,2}, Jing Xia¹, Gao Yuan^{1,2}, Liu Liangyun²

(1.College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China;

2.Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: In order to improve the stability of the small sample data model, a remote sensing detection model of wheat stripe rust with higher accuracy and better robustness was constructed. Firstly, the data of canopy solar-Induced chlorophyll Fluorescence (SIF) were extracted based on radiance and reflectance fluorescence index method, and then combined with reflectance spectral index sensitive to severity of wheat stripe rust, the Gradient Boost Regression Tree (GBRT) was used to detect wheat stripe rust. By comparing GBRT model with CART and Multiple Linear Regression (MLR) model, the results showed that: (1) Reflectivity derivative fluorescence index D705/D722, short-wave infrared Valley reflectance and reflectance ratio fluorescence index R740/R800 were the main factors affecting the accuracy of remote sensing detection of wheat stripe rust. The importance of chlorophyll fluorescence data was higher than that of reflectance spectrum data, and canopy SIF could reflect wheat stripe rust information more sensitively than reflectance spectrum. (2) Compared with CART model and MLR model, the Root Mean Square Error (RMSE) of GBRT model was reduced by 15.50% and 13.49%, and the determination coefficient (R^2) was increased by 6.16% and 11.57% respectively. The estimated DI value of GBRT model is closer to the measured value, and the fluctuation of the estimated result is low, and the robustness of CART model is high. In small sample data, it is easy to divide data sets with different features into subsets of the same feature, and the prediction results fluctuate greatly. The prediction results of MLR model are relatively stable, but its prediction accuracy is low.

Key words: GBRT; Solar-induced chlorophyll fluorescence; Reflectance spectrum; Wheat stripe rust; Disease severity