

基于 LightGBM 的温室番茄冠层 CWSI 预测模型研究

孙 泉¹ 耿 磊² 赵奇慧¹ 杨佳昊² 吕 平³ 李 莉¹

(1. 中国农业大学农业农村部农业信息获取技术重点实验室, 北京 100083;

2. 中国农业大学智慧农业系统集成研究教育部重点实验室, 北京 100083;

3. 浙江施朗龙山工程设计有限公司, 杭州 310000)

摘要: 为研究温室内番茄冠层作物水分胁迫指数(CWSI)问题,通过布设多参数传感器,实时获取温室内外各环境参数。利用灰度关联分析,计算各环境参数与番茄冠层 CWSI 的关联度,根据关联度对环境参数进行排序,同时考虑对模型精度的影响,最终从 9 个环境参数中选取 7 个作为模型输入,建立基于 LightGBM 的温室番茄冠层 CWSI 预测模型。结合贝叶斯算法优化其中的关键参数,将模型预测结果与通过 Jones 经验公式计算出的 CWSI 做相关性分析,在相同的运算环境下,分别与 GBRT 和 SVR 模型对比。试验结果表明,基于贝叶斯优化 LightGBM 模型的决定系数(R^2)、平均绝对误差(MAE)、均方根误差(RMSE)和运算时间分别为 0.960 1、0.021 8、0.031 4 和 0.051 8 s,与 GBRT 和 SVR 模型相比,其 R^2 分别提高 2.14% 和 14.05%,MAE 分别降低 0.009 3 和 0.061 2,RMSE 分别降低 0.009 7 和 0.059 1,时间分别缩短 0.045 9 s 和 0.061 2 s。表明本研究提出的 LightGBM 模型性能更有效地提高了温室番茄冠层 CWSI 的预测精度,为实现温室番茄按需灌溉提供了参考。

关键词: 温室番茄; 作物水分胁迫指数; 预测模型; LightGBM; 热红外

中图分类号: TP181 **文献标识码:** A **文章编号:** 1000-1298(2022)S1-0270-07

CWSI Prediction Model of Greenhouse Tomato Canopy
Based on LightGBM Algorithm

SUN Quan¹ GENG Lei² ZHAO Qihui¹ YANG Jiahao² LÜ Ping³ LI Li¹

(1. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs,
China Agricultural University, Beijing 100083, China

2. Key Laboratory of Smart Agriculture System Integration, Ministry of Education,
China Agricultural University, Beijing 100083, China

3. Zhejiang SLA Engineering Design Co., Ltd., Hangzhou 310000, China)

Abstract: In order to study the prediction of crop water stress index (CWSI) of tomato canopy in greenhouse, through the deployment of multi parameter sensors, the environmental parameter inside and outside the greenhouse can be obtained in real time. Using gray correlation analysis, the correlation degree between environmental parameters and tomato canopy CWSI and the sub factor correlation coefficient between environmental parameters was calculated, the environmental parameters were sorted according to the correlation degree, and the impact on the accuracy of the model was considered. Finally, a total of seven parameters from nine environmental parameters were selected as the model input, and a prediction model of greenhouse tomato canopy crop water stress index (CWSI) based on LightGBM was established. Combined with Bayesian algorithm to optimize the key parameters, the correlation between the prediction results of the model and the CWSI value calculated by Jones empirical formula was analyzed. Under the same computing environment, it was compared with GBRT and SVR models respectively. The experimental results showed that the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE) and operation time of the Bayesian optimized LightGBM model were 0.960 1, 0.021 8, 0.031 4 and 0.051 8 s, respectively. Compared with GBRT and SVR models, R^2 was increased by 2.14% and 14.05% respectively, MAE was reduced by 0.009 3 and

收稿日期: 2022-06-18 修回日期: 2022-08-16

基金项目: 浙江省“尖兵”“领雁”研发攻关计划项目(2022C02013)和国家重点研发计划项目(2019YFD1001903)

作者简介: 孙泉(1996—),男,硕士生,主要从事精细农业系统集成研究,E-mail: sy20213081607@cau.edu.cn

通信作者: 李莉(1978—),女,副教授,博士生导师,主要从事精细农业和农业信息化技术研究,E-mail: lily@cau.edu.cn

0.061 2 respectively, RMSE was reduced by 0.009 7 and 0.059 1 respectively, and the time was shortened by 0.045 9 s and 0.061 2 s respectively. It was showed that the LightGBM model proposed had better performance, which could effectively improve the prediction accuracy of greenhouse tomato canopy CWSI, and provide a strategy for realizing greenhouse tomato on-demand irrigation and a reference for water requirement research.

Key words: greenhouse tomato; CWSI; prediction model; LightGBM; thermal infrared

0 引言

作物水分胁迫指数 (Crop water stress index, CWSI) 是利用作物群体冠层与大气温度差值来反映作物水分胁迫状况的有效指标^[1-3], 可定量研究冠气温差与土壤、植物、大气之间的关系, 构建基于 CWSI 的作物水分状况监测与诊断模型, 对于实现作物水分状况的无损实时监测, 指导精准灌溉具有重要意义^[4]。目前通过热红外成像技术获取冠层叶片温度进而计算 CWSI 来诊断作物水分状况^[5]是相对简单、有效的方法。尚晓英等^[6]利用低空无人机搭载的热红外相机, 获取不同水分处理下的棉花花铃期不同时刻的冠层温度, 计算得到 3 种 CWSI, 验证了作物水分胁迫指数与土壤含水率的幂函数关系。杨帅等^[7]以 4 种水分梯度处理的拔节期玉米为研究对象, 借助无人机可见光和热红外图像, 提取玉米冠层温度信息, 计算 CWSI, 进而反演出玉米地土壤含水率, 判断作物水分胁迫状况。

然而, 上述获得 CWSI 的方法, 需要成本较高的无人机与热像仪, 同时计算方式复杂, 很难在温室中推广使用。随着计算机技术的发展, 使用算法模型^[8]来精确预测 CWSI 成为一种趋势。徐慧慧等^[9]通过测定小白菜的叶片温度、空气温湿度等参数, 确定了叶气温差和空气饱和水汽压差 (Vapor pressure difference, VPD) 的关系, 根据 CWSI 理论计算模式构建了小白菜的 CWSI 模型, 模拟出植物的水分状况。OLUTOBI 等^[10]将空气温湿度、辐射强度、VPD 作为基于数据的黑箱模型 (Data based mechanistic, DBM) 的输入变量, 预测了叶片湿叶、干叶 2 个基准温度, 简化了温室莴苣的 CWSI 计算模型。

本文以温室番茄为研究对象, 获取室内空气温度 (Ta)、室内空气相对湿度 (Rh)、二氧化碳浓度 (CO₂)、基质湿度 (Ms)、基质温度 (Ts)、室外空气温度 (To)、室外相对湿度 (Rho)、光照强度 (LI) 和风速 (Ws) 共 9 个参数, 利用高效梯度提升决策树 (LightGBM) 回归算法^[11], 建立温室番茄 CWSI 的预测模型。并将模型的结果与梯度决策回归树 (Gradient boost regression tree, GBRT) 模型^[12]和支持向量机 (Support vector regression machine, SVR) 模型^[13]的结果进行对比分析。

1 材料与方法

1.1 试验地点与试验设计

试验于 2020 年 11 月在中国农业大学信息与电气工程学院屋顶温室 (40°0′N, 116°21′E) 中进行。温室布置为南北走向, 面积 4.5 m × 3.2 m。试验选取盆栽“粉冠”番茄, 盆容积 2.5 L, 种植容积 2 L。栽培基质为泥炭、蛭石、珍珠岩, 按体积比 3:1:1 均匀混合。施史丹利番茄水溶复合肥“第四元素”, 其中, 40% 为基肥、60% 为追肥^[14], 追肥在开花结果初期和盛期分 2 次施入, 在番茄进入初果期后进行数据采集。试验设置 T1、T2、T3 3 个水分处理, 其基质含水率分别为田间持水量的 65% ~ 75%、55% ~ 65%、30% ~ 55%。每个水分处理 4 株番茄, 共 12 株番茄, 使用基质水分传感器进行监测, 使基质水分分别保持在相应的范围^[15]。

1.2 数据采集

采集数据包括温室内外环境数据与番茄植株冠层热红外图像。使用北京昆仑海岸科技股份有限公司的 JZH-0xx 系列传感器测量温室内空气温度 (Ta)、室内空气相对湿度 (Rh)、二氧化碳浓度 (CO₂)、光照强度 (LI)、基质湿度 (Ms) 和基质温度 (Ts) 6 项参数, 通过 ZigBee 通信网关上传到云平台。同时从气象局获取温室外部气象数据, 分别为室外空气温度 (To)、室外空气相对湿度 (Rho) 和风速 (Ws)。考虑到基质理化性质与一般土壤差异较大, 采用性能稳定的 EC5 传感器作为标准, 进行土壤水分传感器 (KL-TWS-SN 型) 标定, 根据已标定传感器监测的基质含水率确定灌溉量。

番茄冠层图像采集使用菲力尔公司 FLIR T620 型热红外相机, 分辨率 640 像素 × 480 像素, 灵敏度为 0.05℃, 精度 ± 2℃。试验中将便携式计算机 (ThinkPad S3 S440 型) 与热像仪连接, 采用 FLIR Tools 软件拍摄番茄植株冠层热红外图像。番茄初果期植株高度约 50 cm, 为获得较佳的拍摄视野, 相机架设位置始终位于番茄植株上方 1 m 处^[16]。安装位置如图 1 所示。

1.3 番茄冠层 CWSI 计算方法

热红外图像中每个像素点都代表着温度信息, 数据量巨大。为简化计算, 采用 Jones 经验公式^[17]

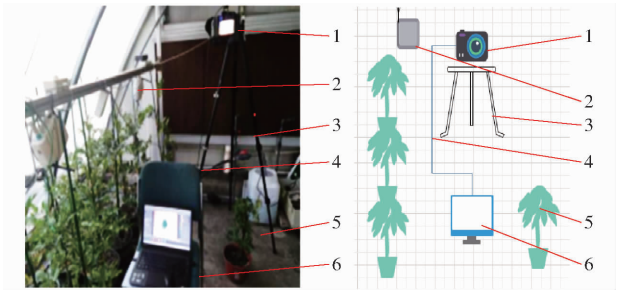


图1 热红外图像采集系统现场图及示意图

Fig.1 Field diagram and schematic of thermal infrared image acquisition system

1. 红外热像仪 2. 温室内环境参数传感器 3. 三脚架 4. 数据线 5. 番茄植株 6. 计算机

来计算 CWSI,使用自然叶片和人工参考面来计算番茄冠层叶片的最低和最高温度,减少风速和太阳辐射变化引起的计算误差,CWSI 的简化计算公式为

$$CWSI = \frac{T_c - T_{wet}}{T_{dry} - T_{wet}} \tag{1}$$

式中 T_c ——冠层实际平均温度,℃
 T_{wet} ——湿参考面温度,℃
 T_{dry} ——干参考面温度,℃

选取 2 片相邻、尺寸相当且完全展开的无病虫害正常生长叶片,一片叶两面均匀涂抹一层凡士林,封闭气孔阻止其蒸腾,作为 CWSI 公式中的自然干参考面温度 T_{dry} ;另一叶片两面喷水并彻底湿润,作为 CWSI 公式中自然湿参考面温度 T_{wet} 。干、湿参考面制成后 1 min,待其温度稳定后进行温度采集^[18]。

实际拍摄的作物冠层热图像中包含土壤、地面等背景信息,需要对作物冠层叶片进行提取。使用红外热像仪配套图像分析软件 FLIR Tools,把一幅热图像中 307 200 个像素点全部导出到 Excel 表格中,对应形成相应温度点。在 Excel 表格中提取 T_{dry} 和 T_{wet} 之间的温度,求取这些温度点平均值作为目标番茄冠层的平均温度 T_c ,以此去除背景信息干扰^[19]。

经过不同水分梯度处理后,于 2020 年 11 月 1—20 日进行数据收集。试验期间每天 09:00—17:00 采集 T_a 、 R_h 、 L_i 、 CO_2 、 M_s 、 T_s 、 To 、 R_h 、 W_s 、 T_c 、 T_{dry} 和 T_{wet} 数据,记录间隔为 30 min,共采集 960 组数据,如图 2 所示。

1.4 红外热像仪误差校准

影响红外热像仪温度测量结果的因素有很多,包括与测量物的距离、环境温湿度、光照、风速等外部因素,还包括热像仪硬件系统温漂的内部因素,通常用黑体来校正^[20]。通过黑体校正,测得温度在 $-0.3 \sim 0.3\text{℃}$ 漂移,符合热像仪的标称精度 $\pm 0.3\text{℃}$,实现对红外测温结果的校准。

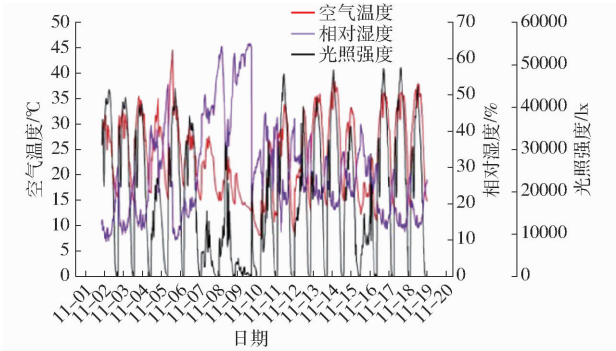


图2 2020 年 11 月 1—20 日部分环境数据曲线

Fig.2 Graphs of partial experimental data from 1st Nov. to 20th Nov. , 2020

1.5 水分胁迫指数预测模型

1.5.1 LightGBM 算法

LightGBM 是由微软团队提出的一种高效梯度提升决策树算法,它是对 GBRT 的一种改进算法,具有更快的训练速度,通过不断优化,让正确结果与预测结果之间的误差不断减小,在不降低预测精度情况下,减少训练时间,同时也大大减少内存的占用。LightGBM 用 2 种方法提升训练速度:①梯度单边采样法 (Gradient based one side sampling, GOSS),GOSS 会从样本中选取有较大梯度的数据从而提升准确率,排除大部分小梯度的样本,仅用剩下的样本计算信息增益。梯度大的实例对信息增益有更大的影响,因此保留梯度大的样本,去掉梯度小的样本。②互斥特征捆绑法 (Exclusive feature bundling, EFB),EFB 可以捆绑互斥特征,使其同时取非零值,即用一个合成特征代替。在实际应用中,虽然特征量比较多,但是由于特征空间十分稀疏,因此 EFB 可以将数据的部分特征进行合并从而降低数据维度^[21]。

1.5.2 基于 LightGBM 的预测模型

基于 LightGBM 的温室番茄水分胁迫指数预测模型,首先将 T_a 、 R_h 、 L_i 、 CO_2 、 M_s 、 T_s 、 To 、 R_h 和 W_s 这 9 个环境因素进行灰度关联性分析,根据相关性分析结果,同时考虑输入变量的选择对预测模型精度的影响,最终选定 T_s 、 R_h 、 CO_2 、 M_s 、 R_h 、 T_a 和 To 这 7 个因素作为 LightGBM 的输入参数,建立环境参数与 CWSI 的回归模型,替代原有的通过获取叶面温度来计算 CWSI 的复杂方法,总体流程见图 3。

1.5.3 灰度关联性分析

灰度关联分析法^[22]的基本思想是量化特征图的几何状态,再计算出参考序列和比较序列之间的关联度。关联度较大时,比较序列的几何发展情况与参考序列更为接近。利用灰色关联理论求出灰色关联度,以此来描述因素间关系的强弱和次序,为合

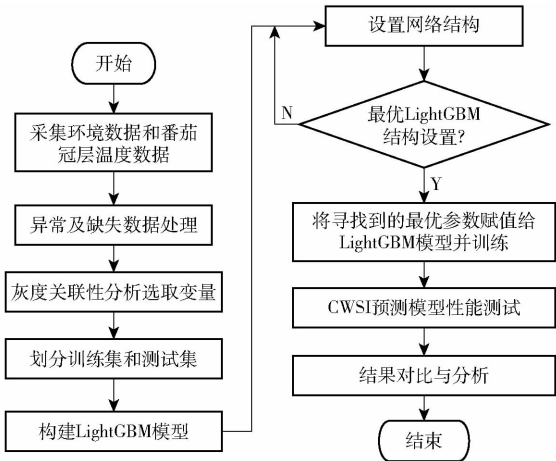


图 3 CWSI 预测模型构建流程图

Fig. 3 Forecast overall flow chart of CWSI

理地选择回归模型的输入因子提供依据。

本文将番茄冠层 CWSI 作为母序列 Y , 将 T_a 、 R_h 、 L_I 、 T_s 、 M_s 、 CO_2 、 To 、 Rho 和 W_s 作为子序列 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 、 X_7 、 X_8 和 X_9 , 建立分析序列

$$Y = (y_1, y_2, \dots, y_j, \dots, y_m) \quad (2)$$
$$(X_1, X_2, \dots, X_n) =$$

$$\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \\ \vdots & \vdots & & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix}_{m \times n} \quad (3)$$

式中 m ——样本数量 n ——子因素数量
 k ——样本组序号
 x_{nm} ——第 m 个样本的第 n 个子因素
对所选样本进行无量纲化处理, 计算式为

$$y_{0j} = \frac{y_j}{y_1} \quad (1 \leq j \leq m) \quad (4)$$

$$x_{0ij} = \frac{x_{ij}}{x_{i1}} \quad (1 \leq i \leq n) \quad (5)$$

关联系数计算式为

$$R = [r_{ij}]_{m \times n} \quad (6)$$

$$r_{ij} = \frac{\min_j \min_i |y_{0j} - x_{ij}| + \rho \max_j \max_i |y_{0j} - x_{ij}|}{|y_{0j} - x_{ij}| + \rho \max_j \max_i |y_{0j} - x_{ij}|} \quad (7)$$

式中 r_{ij} ——关联系数

R ——关联系数的 $m \times n$ 矩阵

ρ ——分辨系数, 一般取 0.5

$\min_j \min_i |y_{0j} - x_{ij}|$ ——2 级最小差

$\max_j \max_i |y_{0j} - x_{ij}|$ ——2 级最大差

各因素关联度计算公式为

$$\varphi_{0j} = \frac{1}{m} \sum_{j=1}^m r_{ij} \quad (1 \leq i \leq n) \quad (8)$$

式中 φ_{0j} ——关联度

1.5.4 评价指标

利用环境参数建立番茄植株冠层 CWSI 预测模型。采用平均绝对误差 (Mean absolute error, MAE)、均方根误差 (Root mean square error, RMSE)、决定系数 (Coefficient of determination, R^2) 评估模型的预测精度^[23]。

2 结果与讨论

2.1 数据集

灌水处理 T1 的 20 d 内基质平均含水率为 48.24% (田间持水量的 75%), 计算出的 CWSI 的平均值为 0.47。灌水处理 T2 的 20 d 内基质平均含水率为 34.73% (田间持水量的 55%), 计算出的 CWSI 的平均值为 0.62。灌水处理 T3 的 20 d 内基质平均含水率为 21.87% (田间持水量的 34%), 计算出的 CWSI 的平均值为 0.85。3 种灌水处理后的 20 d 内基质平均含水率和平均 CWSI 值均存在明显差异, 且传感器测得的基质含水率的数值都大于枯萎含水率^[24]。

试验于 2020 年 11 月采集了 3 种灌溉处理的样本数据共 960 组, 选取同一天 15:00 的热红外图像如图 4 所示。

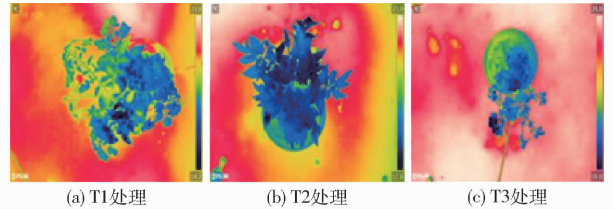


图 4 3 种灌水处理番茄冠层的热红外图像

Fig. 4 Thermal infrared images of tomato canopy under three irrigation treatments

3 种水分处理在相同环境温度下, 基质含水率越低, 番茄冠层温度越高, 各处理温度由高到低依次为 T3、T2、T1, 对应的冠层最低温度分别为 19、17.6、14.2℃。当基质含水率低时, 叶片通过调节气孔开度降低蒸腾速率, 从而减少植株水分散失, 使得番茄植株潜热减少, 显热增加, 叶温升高; 反之当基质含水率高时, 番茄蒸腾速率较高, 冠层叶温较低^[25-27]。

采集到的每组数据中包括 T_a 、 R_h 、 L_I 、 T_s 、 M_s 、 CO_2 、 To 、 Rho 、 W_s 、 T_c 、 T_{dry} 和 T_{wet} 。为保证试验数据划分的随机性和一致性, 将 960 组数据以比例 7:3 随机划分为训练集和测试集, 最终得到训练集共 672 组数据, 用于模型的建立与优化; 测试集共 288 组数据, 用于检验测试模型效果。

2.2 各环境因素与 CWSI 相关性分析

使用灰度关联分析进行特征参数的提取, 并计

算试验数据的 Ta、Rh、LI、Ts、Ms、CO₂、To、Rho 和 Ws 共 9 个因素与温室番茄冠层 CWSI 之间的关联度，如表 1 所示。

表 1 番茄冠层 CWSI 与各因素关联度
Tab.1 Correlation degree between CWSI and various sub-factors in tomato canopy

子序列	因素	与母序列冠层 CWSI 的关联度
X_1	Ta	0.86
X_2	Rh	0.88
X_3	LI	0.77
X_4	Ts	0.92
X_5	Ms	0.89
X_6	CO ₂	0.90
X_7	To	0.82
X_8	Rho	0.91
X_9	Ws	0.75

关联度从大到小顺序为 X_4 、 X_8 、 X_6 、 X_5 、 X_2 、 X_1 、 X_7 、 X_3 、 X_9 ，一般情况下，关联度大于等于 0.8 时，子序列与母序列关联度很好；处于 0.6~0.8 之间关联度一般；小于 0.6 时，表示该子序列与母序列基本不相关。从表 1 可知，Ta、Rh、Ts、Ms、CO₂、To 和 Rho 与母序列冠层 CWSI 的关联度都大于 0.8，表现了这些环境参数与 CWSI 有较好的关联度。各个因素与母序列关联性最好的前 3 位对应的变化范围分别是 Ts (12.3~19.8℃)、Rho (25%~94%)、CO₂ ($3.3 \times 10^{-4} \sim 7 \times 10^{-4} \mu\text{mol/mol}$)。

2.3 预测模型最优参数选择

LightGBM 模型中需优化的参数包括学习率 (learning_rate)、最大深度 (max_depth)、子叶个数 (num_leaves) 和子叶最小数据数 (min_data_in_leaf) 这 4 个参数。采用贝叶斯优化以 5 折交叉验证的平均误差 MSE 作为目标函数，控制迭代次数为 100 次，得到参数优化如表 2 所示。

表 2 LightGBM 模型参数调优结果

参数	范围	寻优值
学习率	(0.1,1)	0.1
最大深度	(4,20)	16
子叶个数	(5,130)	41
子叶最小数据数	(5,30)	24

采用上述建立的 LightGBM 模型对番茄冠层 CWSI 进行预测，将结果与 GBRT 和 SVR 模型的预测值进行对比，后两者模型参数均通过网格搜索和 5 折交叉验证寻优^[28]，SVR 的核函数采用 RBF。SVR 和 GBRT 模型参数调节还包括惩罚系数 (C)、超参数 (γ)、拆分内部节点所需最少样本数 (min_

samples_split)、决策树 (n_estimators)，结果如表 3 所示。

表 3 SVR 和 GBRT 模型参数调优结果
Tab.3 Parameters tuning results of SVR and GBRT models

模型	参数	寻优值
SVR	惩罚系数	23.250 4
	超参数	14.298 0
	最大深度	3
GBRT	子叶最小数据数	2
	拆分内部节点所需最少样本数	2
	学习率	0.01
	决策树	100

2.4 CWSI 预测模型验证

将训练样本集输入已构建的 LightGBM 网络进行训练，然后采用测试样本在已训练的网络进行模型性能测试，将 3 种模型的预测值和根据 Jones 经验公式计算值进行相关性分析。

(1) 模型性能

网络训练时，输入数据集特征分为 4 组，按照灰度关联度排序，由表 1 可以看出，LI 和 Ws 的相关度最低且数值相近，与 To 相差较大。To 与 Ta 数值相差较大。综合考虑输入特征数量分为 4 组，分别是 9、7、6、3，各组包含的具体特征如表 4 所示。模型训练时，分别输入 4 组特征数据，并统计结果，3 种模型性能表现如图 5 所示。

表 4 输入特征分组

Tab.4 Input feature grouping

特征数量	具体特征
9	Ta、Rh、LI、CO ₂ 、Ms、Ts、To、Rho 和 Ws
7	Ts、Rho、CO ₂ 、Ms、Rh、Ta 和 To
6	Ts、Rho、CO ₂ 、Ms、Ta 和 Rh
3	Ts、Rho 和 CO ₂

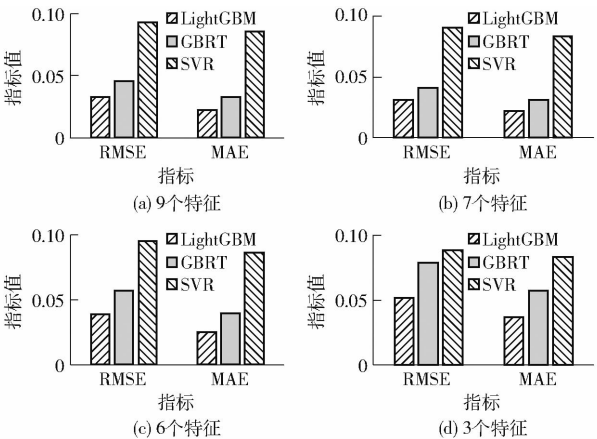


图 5 不同特征数量输入下的模型性能

Fig.5 Model performance under different feature quantity inputs

由图 5 可以看出,随着特征数的减少,模型的 MAE 和 RMSE 会增大,表明输入数据特征越丰富,模型拟合效果越好。3 种模型中,LightGBM 模型的性能表现最好,其 MAE 为 0.021 8 ~ 0.036 6, RMSE 为 0.031 4 ~ 0.051 1,整体性能指标较好,说明该模型的拟合精度较高。从输入特征角度分析,当特征数由 9 个减少到 7 个后,LightGBM 模型的 MAE 变化较小,说明 LI 和 Ws 这 2 个特征对于 CWSI 的影响不大。而当减少的输入特征为 To、LI 和 Ws 时,LightGBM 模型的 MAE 产生了较明显的变化,说明 To 对 CWSI 的影响较大,可为灌溉策略的制定提供参考。GBRT 模型的 MAE 为 0.031 1 ~ 0.056 7, RMSE 为 0.041 1 ~ 0.078 6,其模型受输入特征变化影响较明显,SVR 模型的 MAE 为 0.083 0 ~ 0.082 9, RMSE 为 0.090 5 ~ 0.087 9,其误差随着输入特征的减少而增加,SVR 模型和 GBRT 模型性能均不如 LightGBM 模型。

在保证模型预测精度的前提下,降低模型复杂度,选择输入特征数量为 7,此时 R^2 最高为 0.960 1,而 MAE 和 RMSE 均最低。表 5 为 4 种输入特征数量下,3 种模型的 R^2 。

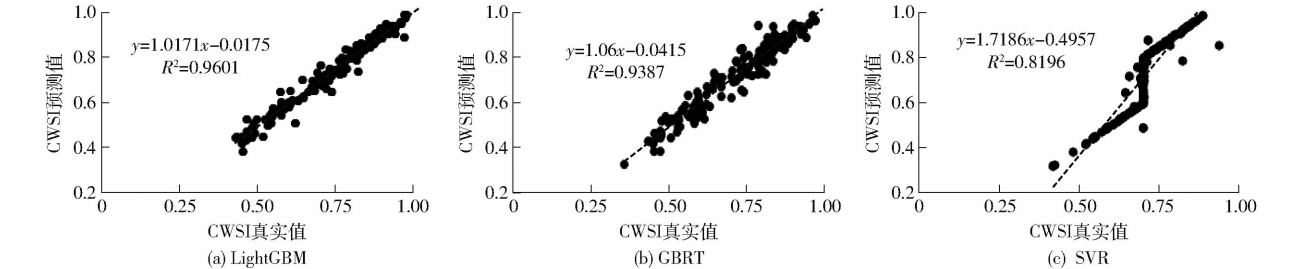


图 6 3 种模型验证结果
Fig. 6 Validation results of three models

3 种模型运行时间和预测精度结果见表 6。可以看出,基于贝叶斯优化 LightGBM 运算时间为 0.051 8 s, GBRT 和 SVR 2 种模型运算时间分别为 0.097 7 s 和 0.113 0 s。原因在于 LightGBM 模型训练过程使用 GOSS 方法,可以减少大量只具有小梯度的数据实例,在计算信息增益时可减少时间和空间上的开销,运算速度明显更快。

LightGBM 模型和 GBRT 模型对弱决策回归树的要求不高,在低维度情况下,数据规模对模型影响

表 6 模型预测精度和运行时间

Tab. 6 Result of model prediction accuracy and running time				
模型	RMSE	MAE	R^2	t/s
LightGBM	0.031 4	0.021 8	0.960 1	0.051 8
GBRT	0.041 1	0.031 1	0.938 7	0.097 7
SVR	0.090 5	0.083 0	0.819 6	0.113 0

表 5 3 种模型在测试集的 R^2
Tab. 5 R^2 for three models

特征数量	LightGBM	GBRT	SVR
9	0.959 1	0.915 5	0.816 1
7	0.960 1	0.938 7	0.819 6
6	0.936 0	0.876 3	0.803 2
3	0.892 9	0.741 6	0.854 4

(2)模型预测验证

当输入特征数量为 7 时,为进一步说明 LightGBM 模型在番茄冠层 CWSI 预测中的优越性,同时利用 GBRT 和 SVR 模型对番茄冠层 CWSI 进行预测并对比。GBRT 和 SVR 模型的预测值与 Jones 经验公式计算值的 R^2 分别为 0.938 7 和 0.819 6,而由 LightGBM 模型得到 R^2 为 0.960 1,与 GBRT 和 SVR 模型结果相比,其 R^2 分别提高 2.14% 和 14.05%,因 R^2 越小,表示模型拟合效果越差,不足以捕获数据中的变化规律。因此,GBRT 和 SVR 模型的预测效果均劣于 LightGBM 模型。可以看出基于贝叶斯优化的 LightGBM 模型的整体预测精度较为理想,该模型可以很好地拟合环境因子与番茄冠层 CWSI 之间复杂的非线性关系。3 种模型验证结果如图 6 所示。

不大,所以对特征空间的异常具有很好的鲁棒性。同时,二者在建模过程中数据不需要归一化,特征的作用只是用来分裂节点,叶节点的值与特征值无关,是否归一化并不影响叶节点值和梯度提升的进程^[29]。LightGBM 模型是 GBRT 模型的优化,采用了 GOSS 和 EFB 方法,使得精度和速度得到进一步改善。

试验结果表明,基于 LightGBM 的温室番茄冠层 CWSI 预测模型可以提取数据的高层次特征,获得较传统方法更好的非线性函数逼近效果,在一定程度上提高了算法的精度和泛化能力,在农业大数据挖掘中能够表现出良好的特征学习能力。

3 结论

(1)利用灰度关联算法筛选出与温室番茄冠层 CWSI 关联度大的 Ta、Rh、CO₂、Ms、Ts、To、Rho 7 个特征参数作为输入,减少了预测模型输入变量的个

数,降低了复杂度,提高了模型预测精度。

(2)提出了一种基于温室环境参数的温室番茄冠层 CWSI 预测方法。经试验表明,当输入特征数量为 7 时,基于贝叶斯优化 LightGBM 模型在测试集中的 R^2 、MAE、RMSE 分别为 0.960 1、0.021 8 和 0.031 4。与 GBRT 和 SVR 模型进行对比,其 R^2 分

别提高 2.14% 和 14.05%,MAE 分别降低 0.009 3 和 0.061 2,RMSE 分别降低 0.009 7 和 0.059 1,时间分别缩短了 0.045 9 s 和 0.061 2 s。表明 LightGBM 模型能够有效提高温室番茄冠层 CWSI 的预测精度,验证了基于贝叶斯优化 LightGBM 模型在预测温室番茄冠层 CWSI 的有效性。

参 考 文 献

- [1] JÉFFERSON D O C, COELHO R D, TIMÓTEO H D S B, et al. Canopy thermal response to water deficit of coffee plants under drip irrigation[J]. *Irrigation and Drainage*, 2020, 69(3):472–482.
- [2] FITRIYAH A, FATIKHUNNADA A, OKURA F, et al. Analysis of the drought mitigated mechanism in terraced paddy fields using CWSI and TVDI indices and hydrological monitoring[J]. *Sustainability*, 2019, 11(24):6897.
- [3] KING B A, SHELLIE K C, TARKALSON D D, et al. Data-driven models for canopy temperature-based irrigation scheduling[J]. *Transactions of the ASABE*, 2020, 63(5):1579–1592.
- [4] 张立元,牛亚晓,韩文霆,等.大田玉米水分胁迫指数经验模型建立方法[J].*农业机械学报*, 2018, 49(5):233–239. ZHANG Liyuan, NIU Yaxiao, HAN Wenting, et al. Establishing method of crop water stress index empirical model of field maize[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(5):233–239. (in Chinese)
- [5] 刘又夫,周志艳,田麓弘,等.红外热成像技术在农业中的应用[J].*农业工程*, 2019, 9(11):102–110. LIU Youfu, ZHOU Zhiyan, TIAN Luhong, et al. Application of infrared thermal imaging technology in agriculture[J]. *Agricultural Engineering*, 2019, 9(11):102–110. (in Chinese)
- [6] 尚晓英,张智韬,边江,等.基于无人机热红外的水分胁迫指数与土壤含水率关系研究[J].*节水灌溉*, 2019(4):16–21. SHANG Xiaoying, ZHANG Zhitao, BIAN Jiang, et al. Study on the relationship between water stress index and soil moisture content based on UAV thermal infrared[J]. *Water Saving Irrigation*, 2019(4):16–21. (in Chinese)
- [7] 杨帅,陈俊英,周永财,等.无人机热红外遥感反演玉米根域土壤含水率方法研究[J].*节水灌溉*, 2021(3):12–18. YANG Shuai, CHEN Junying, ZHOU Yongcai, et al. A study on the method of UAV thermal infrared remote sensing to retrieve soil moisture content in corn root zone[J]. *Water Saving Irrigation*, 2021(3):12–18. (in Chinese)
- [8] 杨明欣,高鹏,陈文彬,等.基于机器学习的油青菜心水分胁迫研究[J].*华南农业大学学报*, 2021, 42(5):117–126. YANG Mingxin, GAO Peng, CHEN Wenbin, et al. Research of *Brassica chinensis* var. *parachinensis* under water stress based on machine learning[J]. *Journal of South China Agricultural University*, 2021, 42(5):117–126. (in Chinese)
- [9] 徐慧慧,张亚琪,于立尧,等.小白菜冠层温度分析与 CWSI 模型构建[J].*上海交通大学学报(农业科学版)*, 2018, 36(3):15–20, 44. XU Huihui, ZHANG Yaqi, YU Liyao, et al. Pakchoi canopy temperature analysis and CWSI model construction[J]. *Journal of Shanghai Jiaotong University(Agricultural Science)*, 2018, 36(3):15–20, 44. (in Chinese)
- [10] OLUTOBI A, IVAN G, SVEN P, et al. Dynamic modelling of the baseline temperatures for computation of the crop water stress index (CWSI) of a greenhouse cultivated lettuce crop[J]. *Computers and Electronics in Agriculture*, 2018, 153:102–114.
- [11] KE G, MENG Q, FINLEY T, et al. Light GBM: a highly efficient gradient boosting decision tree[C]//*Neural Information Processing Systems Foundation*, 2017:3149–3157.
- [12] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *Annals of Statistics*, 2001, 29(5):1189–1232.
- [13] CORTES C, VAPNIK V. Support vector network[J]. *Machine Learning*, 1995, 20(3):273–297.
- [14] 岳霆,杨平,巩法江,等.水分对设施番茄果实发育和产量的影响[J].*农业科技通讯*, 2017(7):212–214. YUE Ting, YANG Ping, GONG Fajiang, et al. Effects of water on fruit development and yield of tomato in greenhouse[J]. *Bulletin of Agricultural Science and Technology*, 2017(7):212–214. (in Chinese)
- [15] 刘朝霞.土壤干旱胁迫对番茄根系生长、气孔特性及保护酶活性的影响[D].南京:南京信息工程大学,2016. LIU Chaoxia. Effects of soil drought stress on root growth, stomatal characteristics and antioxidant enzyme of tomato crops[D]. Nanjing: Nanjing University of Information Science & Technology, 2016. (in Chinese)
- [16] 马丽,张旭东,邢子正,等.基于骨架扫描策略的生猪耳根体表温度 FDSST 检测方法[J].*农业机械学报*, 2020, 51(增刊1):371–377. MA Li, ZHANG Xudong, XING Zizheng, et al. FDSST detection method for pig ear base surface temperature based on skeleton scanning strategy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(Supp. 1):371–377. (in Chinese)
- [17] JONES H G, STOLL M, SANTOS T, et al. Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine[J]. *Journal of Experimental Botany*, 2002, 53(378):2249–2260.
- [18] 吴晓磊.基于红外热成像的棉花水分状况诊断研究[D].北京:中国农业科学院,2016. WU Xiaolei. Diagnosis research of cotton water status based on infrared thermal imaging[D]. Beijing: Chinese Academy of Agricultural Sciences, 2016. (in Chinese)
- [19] PADHI J, MISRA R K, PAYERO J O. Estimation of soil water deficit in an irrigated cotton field with infrared thermography[J]. *Field Crops Research*, 2012, 126(1):45–55.

- [52] 肖慧,孙柯,屠康,等. 便携式葡萄专用可见-近红外光谱检测仪器开发与实验[J]. 食品科学,2019,40(8):300-305.
XIAO Hui, SUN Ke, TU Kang, et al. Development and application of a specialized portable visible and near-infrared instrument for grape quality detection[J]. Food Science, 2019, 40(8): 300-305. (in Chinese)
- [53] 张宁,赵旭,兰义宾,等. 中国东亚种红葡萄酒的颜色特征及酚类组成研究[J]. 中国酿造,2022,41(5):34-41.
ZHANG Ning, ZHAO Xu, LAN Yibin, et al. Color features and phenolic composition of red wines made from East Asian Vitis species native to China[J]. China Brewing, 2022, 41(5): 34-41. (in Chinese)
- [54] 蔡正云,吴龙国,王菁,等. 宁夏赤霞珠葡萄水分含量的高光谱无损检测研究[J]. 食品工业科技,2017,38(2):79-83.
CAI Zhengyun, WU Longguo, WANG Jing, et al. Non-destructive determination of moisture composition in Ningxia wine grapes based on visible near-infrared hyperspectral imaging technique[J]. Science and Technology of Food Industry, 2017, 38(2): 79-83. (in Chinese)
- [55] 高升,王巧华. 基于可见/近红外透射光谱技术的红提糖度和含水率无损检测[J]. 中国光学,2021,14(3):566-577.
GAO Sheng, WANG Qiaohua. Non-destructive testing of red globe grape sugar content and moisture content based on visible/near infrared spectroscopy transmission technology[J]. Chinese Optics, 2021, 14(3): 566-577. (in Chinese)
- [56] 陈辰,鲁晓翔,张鹏,等. 红提葡萄 VC 含量的可见/近红外检测模型[J]. 食品与机械,2015,31(5):70-74.
CHEN Chen, LU Xiaoxiang, ZHANG Peng, et al. Determination of vitamin C in red globe grape based on visible/near-infrared diffuse reflectance spectroscopy[J]. Food and Machinery, 2015, 31(5): 70-74. (in Chinese)
- [57] 高升,王巧华,李庆旭,等. 基于近红外光谱的红提维生素 C 含量、糖度及总酸含量无损检测方法[J]. 分析化学,2019,47(6):941-949.
GAO Sheng, WANG Qiaohua, LI Qingxu, et al. Non-destructive detection of vitamin C, sugar content and total acidity of red globe grape based on near-infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2019, 47(6): 941-949. (in Chinese)
- [58] 罗一甲,祝赫,李潇涵,等. 赤霞珠酿酒葡萄总酚含量的近红外光谱定量分析[J]. 光谱学与光谱分析,2021,41(7):2036-2042.
LUO Yijia, ZHU He, LI Xiaohan, et al. Quantitative analysis of total phenol content in cabernet sauvignon grape based on near-infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2021, 41(7): 2036-2042. (in Chinese)
- [59] XIAO H, FENG L, SONG D, et al. Grading and sorting of grape berries using visible-near infrared spectroscopy on the basis of multiple inner quality parameters[J]. Sensors, 2019, 19(11): 2600.

(上接第 276 页)

- [20] 马丽. 基于热红外视频的生猪耳根体表温度自动检测技术研究[D]. 北京:中国农业大学,2020.
MA Li. Research on automatic detection technology of pig ear-root body surface temperature based on thermal infrared video [D]. Beijing:China Agricultural University, 2020. (in Chinese)
- [21] CHEN Y, HU M, XIE Y. Claim frequency predicting based on LightGBM[J]. Journal of Nonlinear and Convex Analysis, 2020, 21(8):1759-1770.
- [22] 刘梦雅,毛剑琳. 基于灰度关联分析法的卷积神经网络算法优化[J]. 电子科技, 2018, 31(6): 84-88, 95.
LIU Mengya, MAO Jianlin. Optimization of convolutional neural network algorithm based on gray relational analysis[J]. Electronic Science and Technology, 2018, 31(6): 84-88, 95. (in Chinese)
- [23] LI Li, CHEN Shiwan, YANG Chengfei, et al. Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm[J]. Journal of Cleaner Production, 2020, 261: 121-136.
- [24] 赵奇慧,李莉,张森,等. 基于迁移学习的温室番茄叶片水分胁迫诊断方法[J]. 农业机械学报, 2020, 51(增刊1): 340-347, 356.
ZHAO Qihui, LI Li, ZHANG Miao, et al. Water stress diagnosis algorithm of greenhouse tomato based on fine-tuning learning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp.1): 340-347, 356. (in Chinese)
- [25] 国银银. 基于茎秆直径微变化和叶温的温室葡萄水分亏缺诊断与蒸腾模拟[D]. 杨凌:西北农林科技大学, 2019.
GUO Yinyin. Water deficit diagnosis and transpiration simulation of greenhouse grape based on stem diameter micro change and leaf temperature[D]. Yangling: Northwest A&F University, 2019. (in Chinese)
- [26] SILVA B B D, RAO T V R. The CWSI variations of a cotton crop in a semi-arid region of Northeast Brazil[J]. Journal of Arid Environments, 2005, 62(4):649-659.
- [27] MUSTAFA Ü, RIZA K, BURÇAK K, et al. The crop water stress index (CWSI) for drip irrigated cotton in a semi-arid region of Turkey[J]. African Journal of Biotechnology, 2011, 10(12): 2258-2273.
- [28] ZHANG T, SU H, YANG X, et al. Remote sensing prediction of global subsurface thermohaline based on LightGBM[J]. Journal of Remote Sensing, 2020, 24(10):1255-1269.
- [29] SHARMA A, SINGH B. AE-LGBM: sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and LightGBM[J]. Computers in Biology and Medicine, 2020, 125:103964.