

基于 NIR 光谱和梯度提升决策树的农作物含水率快速检测平台研究

杨 宁¹ 方 啸¹ 程 巍¹ 张钊源¹ 王亚飞² 陈 思³ 毛罕平²

(1. 江苏大学电气信息工程学院, 镇江 212013; 2. 江苏大学农业工程学院, 镇江 212013;
3. 江苏大学流体机械工程技术研究中心, 镇江 212013)

摘要: 农作物含水率与其质量息息相关,含水率过高易引发霉变,过低则会影响形态完整,然而现有便携检测方法存在精度低、检测速度慢等问题,烘干法处理单样本需 40 min,定容称量法检测误差为 1.2%,为满足农作物含水率快速检测需求,本文设计了一种基于近红外 (Near-infrared, NIR) 光谱和梯度提升决策树 (Gradient boosting decision tree, GBDT) 的农作物含水率快速检测平台。基于 Zemax 优化设计了离轴切尼-特纳 (Czerny - Turner, CT) 光学结构,采用 ZYNQ 芯片内嵌线阵 CCD 解码算法并与手机连接,在手机上建立萨维茨基-戈雷 (Savitzky - Golay, SG) 去噪和分段直接标准化 (Piecewise direct standardization, PDS) 数据迁移算法,通过光谱采样平台采样后再加载手机的 GBDT 模型预测含水率。Zemax 仿真证明离轴设计光学结构在体积和像差控制上均优于传统球面设计,在像面边缘 950 nm 处仍可以保持 2 nm 分辨率。SG 去噪可使信噪比 (SNR) 从 18.11 dB 提升到 35.10 dB,基于 GBDT 建立的含水率预测模型在验证集回归预测中决定系数 R^2 和预测均方根误差 (RMSEP) 分别为 0.877 7 和 0.135 4%,在测试集中 R^2 和 RMSEP 分别为 0.861 7 和 0.121 4%,含水率绝对误差小于等于 0.4%。检测平台具有体积小、速度快、方便携带等优势,单样本检测时间相较于烘干法 (40 min) 小,仅需 0.5 min,满足农作物含水率快速高效检测需求。

关键词: 含水率; 快速检测; 近红外光谱; 切尼-特纳结构; 梯度提升决策树

中图分类号: S237 文献标识码: A 文章编号: 1000-1298(2026)08-0299-09 **OSID:** 

Platform for Rapid Detection of Crop Moisture Content Based on NIR Spectroscopy and GBDT

YANG Ning¹ FANG Xiao¹ CHENG Wei¹ ZHANG Zhaoyuan¹ WANG Yafei² CHEN Si³ MAO Hanping²

(1. School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

2. School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China

3. Research Center of Fluid Machinery Engineering and Technology, Jiangsu University, Zhenjiang 212013, China)

Abstract: Water content detection is essential for preventing mold growth and ensuring the proper storage and transportation of crops. However, existing portable detection methods often suffer from low accuracy and efficiency. To meet the demand for rapid crop water content measurement, a fast detection platform was proposed based on near-infrared (NIR) spectroscopy and a gradient boosting decision tree (GBDT) model. An off-axis Czerny - Turner (CT) optical structure was optimized by using Zemax, and a ZYNQ chip was employed to implement embedded decoding algorithms for a linear array CCD, with the system connected to a mobile phone. Savitzky - Golay (SG) smoothing and a piecewise direct standardization (PDS) data transfer algorithm were implemented on the mobile device. After spectral acquisition, the GBDT model running on the mobile phone was used to predict crop water content. Zemax simulations demonstrated that the off-axis optical design outperformed traditional spherical designs in terms of volume reduction and aberration control, maintaining a spectral resolution of 2 nm at the 950 nm edge of the image plane. SG smoothing improved the signal-to-noise ratio (SNR) from 18.11 dB to 35.10 dB. The GBDT-based prediction model achieved a coefficient of determination (R^2) of 0.877 7 and a root mean

收稿日期: 2025-01-20 修回日期: 2025-03-18

基金项目: 国家重点研发计划青年科学家项目(2022YFD2000200)和国家自然科学基金面上项目(32171895)

作者简介: 杨宁(1984—),男,教授,博士生导师,主要从事智慧农业作物病害预警研究,E-mail: yangning7410@163.com

通信作者: 毛罕平(1961—),男,教授,博士生导师,主要从事智能化农业装备技术研究,E-mail: maohp@ujs.edu.cn

square error of prediction (RMSEP) of 0.135 4% on the validation set. On the test set, the model attained an R^2 of 0.861 7 and an RMSEP of 0.121 4%, with prediction absolute errors within 0.4%, meeting the requirements for rapid and efficient crop water content detection.

Key words: moisture content; rapid detection; near-infrared spectroscopy; Czerny – Turner structure; gradient boosting decision trees

0 引言

农作物含水率过高可能导致产品变质、发霉,甚至引发食品安全问题^[1],而含水率过低则会使农作物变得易碎,对于种子过低的含水率则会影响萌发率^[2-3]。例如,GB 1353—2018《玉米》规定 3 等玉米含水率小于等于 14.0%,不完善粒含量小于等于 8.0%。因此,对农作物含水率进行检测至关重要。

农作物含水率检测技术主要有烘干法、电化学法、核磁共振法(Nuclear magnetic resonance, NMR)、光谱法等。传统烘干法具有很高精度,但需将样本置于干燥箱干燥数小时^[4],不符合实时快速检测发展需求。使用卤素含水率测定仪干燥样本可将含水率检测时间缩减到 1 h 以内且保证较高精度,但每台设备只能处理 1 份样本,无法实现大量样本的快速检测^[5]。电化学法包括阻抗法和介电系数法,阻抗法是一种成熟的测量技术,通过测量极间阻抗分析含水率^[6-7],具有测量速度快、无损检测的特点,因此广泛应用于生产活动中;介电系数法则是测量高频激励下的极间电容推算介电系数,再根据介电系数建模分析含水率^[8-10]。但目前电化学法均要求样本和电极间紧密接触,其接触面积越大扰动越小,因此电化学法通常要求将长探针完全插入待测样本中以获得稳定的检测结果,对于少量样本难以精确检测。

随着对精度要求的不断提高,NMR 法和光谱法近年发展迅速,ZHANG 等^[11]利用低场脉冲核磁共振(LF-NMR)检测农作物多种组分,例如含水率和脂肪含量,但 NMR 设备造价昂贵、体积巨大、操作复杂,难以大范围推广。光谱法利用近红外光(Near-infrared, NIR)波段的漫反射吸收光谱分析含水率,水分子在 NIR 照射下发生分子振动,消耗光能产生吸收带^[12],光谱法通常将光谱仪、积分球、卤素光源组合为检测系统。通过 NIR 光谱传感器采集光谱数据,再用各种算法训练模型,实现对样本分析目的。MOON 等^[13]将光谱数据通过卷积神经网络(CNN)进行训练,可以对食品新鲜度进行预测。刘文政等^[14]基于偏最小二乘回归(PLSR)、支持向量机回归(SVR)等算法对采集的光谱数据进行模型训练,实现对葡萄品质预测。彭彦昆等^[15]基于 PLSR 建立近红外光谱和猪肉含水率检测模型,实现

了含水率实时检测和猪肉分级。在发展应用中,光谱法存在设备成本高、体积大,光谱仪成像效果差的问题。张静等^[16]使用便携式 NIR 光谱仪代替传统的傅里叶变换光谱仪检测水稻含水率,完成了初步的小型化系统。为了进一步提高集成度,李扬裕等^[17]在光栅光谱仪的基础上采用平板波导将光限定在一个扁平结构里,可有效降低光谱仪垂直厚度。GATTINGER 等^[18]使用 DMD 数字微镜元件和单像素光电管的切尼-特纳(Czerny – Turner, CT)结构代替线阵探测器,使其能在 NIR 波段实现较高分辨率。JEON 等^[19]使用 MEMS 技术将活动光栅微型化并将 CT 结构的准直镜和聚焦镜融合,同样配合单像素光电管,体积更小。但这种以单像素光电管配合活动部件的传感器难以在分辨率和光谱范围间平衡,而且光谱中心分辨率和边缘分辨率相差较大,存在较大像差。除了 CT 结构还有使用光学玻璃制成透镜的透射式光谱仪,该结构光学元件可以线性排列,因此可以很好地控制体积^[20],但 NIR 能量较低,在透过光学玻璃时存在较大衰减,多用于可见光。

为解决现有含水率快速检测系统精度低、检测速度慢、体积大等问题,进一步提升系统集成度、操作便捷性,本文设计一种高分辨率高集成度的物联网快速检测平台,将 NIR 传感器、光源和旋转载物平台结合,采集光谱数据进行预处理并建立含水率预测模型。以实现无需预处理样本的现场实时检测,即在手机上进行模型推理分析预测样本含水率。

1 传感器原理和结构

1.1 光学设计

为兼顾波长范围和分辨率,NIR 传感器设计为 M 形 CT 结构。安装光纤适配器可以螺纹连接光纤,入射光由光纤射出照亮狭缝形成物面,经过反射准直镜准直为平行光并反射到刻线光栅上,光栅分光后的单色平行光反射到聚焦镜上,由聚焦镜将各个单色光聚焦在 CCD 探测器上形成像面。由于光栅分光后不同波长单色光的衍射角不同,经过聚焦后落在像面上的位置也不同。

在 Zemax 导入参数并以序列模式进行设计,设计光谱探测波段为 950 ~ 1 700 nm,数值孔径为 0.09,狭缝宽度为 25 μm,光栅线数为 100 gr/mm 的

反射式光栅,利用光栅的-1衍射级,准直镜和聚焦镜表面镀银。图 1a 为采用离轴设计的光学结构,具有非常紧凑的光路,即使不进行光路交叉也能很好地控制体积。图 1b 为球面设计的光学结构,对比离轴设计,光路更长体积更大。相较于球面设计,离轴设计光学结构投影面积缩小 37.2%。

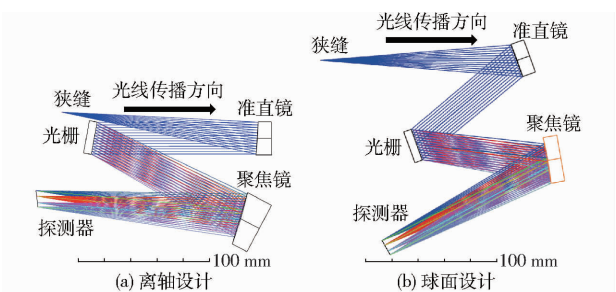


图 1 光路的三维布局图

Fig. 1 Three-dimensional layout diagrams of optical path

相较于交叉型 CT 结构,M 形 CT 结构在像差控制上具有明显优势,其在像面边缘依旧可以保持较高分辨率^[21]。在 Zemax 物面设置波段 950 ~ 1 700 nm 的点光源,由于选型的 CCD 探测器感光窗口尺寸为 12 800 μm ,因此在 Zemax 中也关注同一尺寸的像面。图 2a 为离轴设计像面全视场点列图,观察到 12 800 μm $\times$ 12 800 μm 像面可以完整覆盖 950 ~ 1 700 nm 光谱,分光波长 950 nm 和 952 nm 靠近像面左侧边沿,分光波长 1 698 nm 和 1 700 nm 靠近像面右侧边沿。图 2a 左侧局部放大图为分光波长 950 nm 和 952 nm 处的全视场像面点列图,可以看到在波长间隔为 2 nm 时仍有分辨能力。图 2a 靠近中心的局部放大图为离轴设计分光波长 1 200 nm 和 1 202 nm 处的全视场像面点列图,综合图 2a 的局部放大图可发现边缘画质没有明显损失,在整个目标波段内都能保持 2 nm 分辨率。若选择更窄的狭缝可以获得更高分辨率,但过窄狭缝会影响通光量,所以在保证分辨率的情况下应尽可能选择宽狭缝。图 2b 为球面设计像面全视场点列图,像面同样可以完整覆盖 950 ~ 1 700 nm 光谱,但观察图 2b 2 个局部放大图,即波长 950 nm 和 952 nm 处以及 1 200 nm 和 1 202 nm 处的像面全视场点列图,可以发现分辨率明显下降,尤其是在像场边缘波长 950 nm 处受彗差影响分辨率不足 2 nm,而靠近中心位置波长 1 200 nm 处也存在明显像散。这是由于球面镜无法矫正像差,多块反射镜像差叠加的结果。

离轴设计的分辨率优势同样反映在惠更斯点扩散函数(Point spread function, PSF)上。图 3 为离轴设计波长 1 500 nm 和 1 502 nm 的 PSF,观察到中心斑为长条形且为线状延展,强度在中心最高边缘迅速衰减且没有重合,满足分辨率 2 nm 需求。

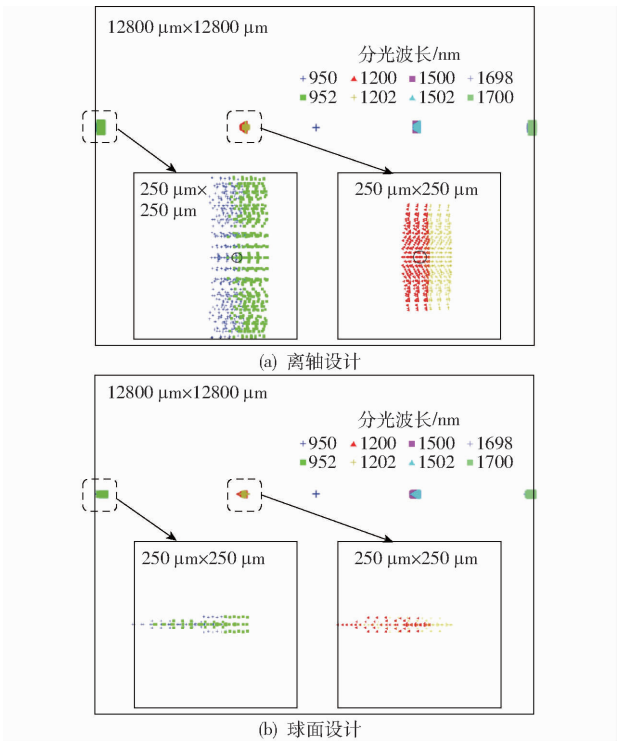


图 2 像面全视场点列图

Fig. 2 Point diagrams of full FoV of image plane

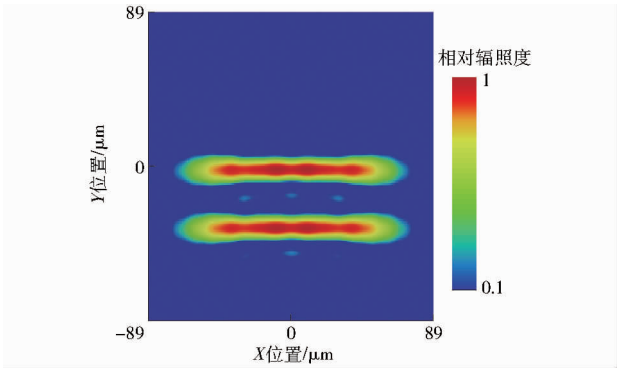


图 3 复色光对数惠更斯 PSF

Fig. 3 Polychromatic Huygens PSF (log scale)

调制传递函数 (Modulation transfer function, MTF)描述光学系统对空间频率的响应能力,即系统保持图像细节的能力。良好的光学系统能更好地分辨和传输不同波长光信号,体现在 MTF 中表现为在高频区域具有较高的 OTF 模值且具有线性的衰减特性,非线性衰减意味着存在成像失真。图 4 为离轴设计光学系统的 MTF,在中低空间频率时,所有视场的子午方向和弧矢方向都呈现线性均匀衰减,高空间频率时保持为周期性衰减。综上所述,离轴设计的光学系统具有很强的边缘像差抑制能力,并且在整个像面都具有很好的分辨率,因此采用此设计。

1.2 电路设计

电路部分如图 5 所示,ZYNQ XC7Z020 芯片产生时钟信号和复位信号驱动线阵传感器,LT1818 作

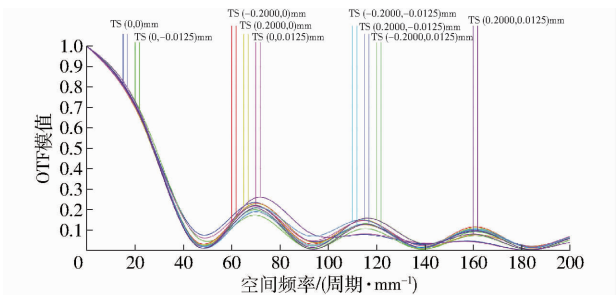


图 4 全视场复色光惠更斯 MTF

Fig. 4 Polychromatic full FoV Huygens MTF

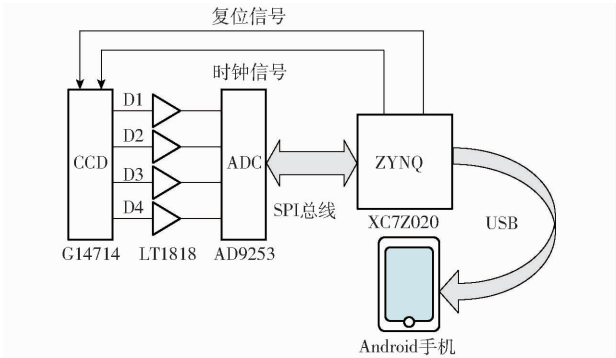


图 5 驱动电路原理图

Fig. 5 Drive circuit schematic

为缓冲器。AD9253 作为 ADC 对 G14714 - 1024DG 的每个像素进行采样,在 ZYNQ 的 PL 部分完成积分时间控制和 ADC 交互数据读取,PS 部分通过 USB 将光谱原始数据发送到手机。NIR 传感器内部贴消光绒以减小杂散光反射干扰。

1.3 NIR 传感器结构

图 6 为 NIR 传感器结构图,狭缝尺寸为 $400\text{ }\mu\text{m}\times 25\text{ }\mu\text{m}$,表面发黑处理防止产生杂散光。使用 6061 - T6 镀银作为准直反射镜和聚焦反射镜,镜面为离轴设计。分光光栅为 100 gr/mm 的平面衍射光栅,Pyrex 玻璃作为基板表面镀铝并刻线制成,光学面尺寸为 $25\text{ mm}\times 25\text{ mm}$ 。选用滨松 G14714 - 1024DG 线阵传感器作为光探测器,具有 InGaAs 光电二极管阵列,对波段 $950\sim 1\text{ }700\text{ nm}$ 光具有有效感光能力,窗口长度为 $12\text{ }800\text{ }\mu\text{m}$ 。这款 NIR 传感器对比常用的光谱仪具有明显成本优势。由于镜片尺寸小巧,因此加工成本和加工难度都大幅下降。

2 农作物光谱-含水率模型建立

以玉米籽粒含水率快速检测为例,使用 Eigenvector Research, Inc. (EVRI)提供的玉米 NIR 光谱训练含水率回归预测模型,再用 NIR 传感器结合手机对 30 个制备的郑单 958 玉米样本扫描光谱并作为测试集,在本地回归预测含水率。由于训练集光谱波段为 $1\text{ }100\sim 1\text{ }700\text{ nm}$,NIR 传感器采集的光谱测试集光谱同样截取波段 $1\text{ }100\sim 1\text{ }700\text{ nm}$ 。

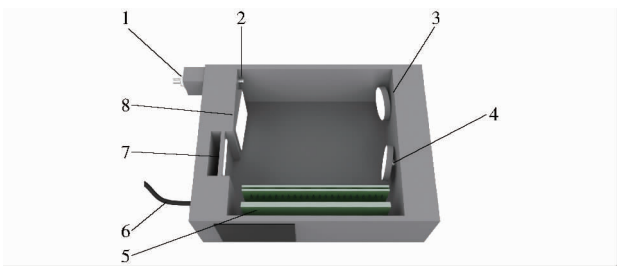


图 6 NIR 传感器结构

Fig. 6 NIR sensor structure

1. SMA905 型光纤耦合器 2. 狭缝 3. 准直镜 4. 聚焦镜 5. 驱动电路 6. USB 数据线 7. 探测器 8. 光栅

2.1 光谱去噪

由于从光源到样品到数据采集会引入光噪声、环境噪声、电子噪声等干扰,降噪时需注意细节损失对模型迁移产生有害影响,分别采用小波变换去噪、萨维茨基-戈雷 (Savitzky - Golay, SG) 去噪、高斯去噪进行预处理。小波变换去噪是一种利用小波变换分解重构和阈值处理进行降噪的方法,具体处理方法有默认阈值去噪处理、给定阈值去噪处理、强制去噪等。SG 去噪通过拟合多项式到数据的 1 个子集 (滑动窗口内数据点)来平滑数据,该方法可有效保留信号高频特征,如峰值、尖锐转折点等有重要局部特征信号。高斯去噪使用高斯函数作为权重系数,对信号进行加权平均以平滑数据,高斯权重使得中心数据点有更高权重,能有效保留信号局部特征。对 EVRI 提供的数据添加噪声并分别使用 3 种去噪方法,计算信噪比 (SNR)、峰值信噪比 (PSNR) 为结果进行对比,如表 1 所示。

表 1 3 种去噪方法结果对比

Tab. 1 Comparison of three denoising results dB

去噪方法	SNR	PSNR
小波去噪	26. 23	31. 22
SG 去噪	35. 10	39. 94
高斯去噪	32. 07	36. 97
含噪数据	18. 11	24. 01

由表 1 可知,SG 去噪效果最优,能很好地还原原始数据。

2.2 异常样本剔除

为提高训练集质量,使用蒙特卡洛交叉验证 (Monte Carlo cross validation, MCCV)处理训练集以避免离散样本误差过大影响训练集,在 80 个样本中保留 70 个作为训练集。图 7 为 MCCV 处理后的每个样本吸光度平均值和标准差,将平均值和标准差离散的 10 个样本剔除,取红框内的样本作为训练样本。

2.3 数据迁移

为提高不同设备间模型的通用性,可使用直接

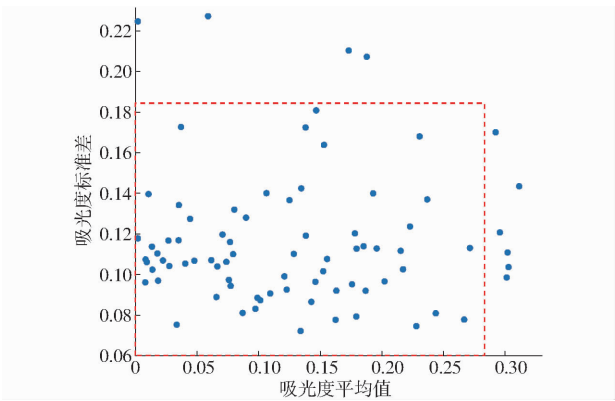


图 7 MCCV 处理结果

Fig. 7 MCCV processing results

标准化(Direct standardization, DS)或分段直接标准化(Piecewise direct standardization, PDS)对数据进行迁移^[22-23]。图 8 为 DS 和 PDS 流程图,2 种方法均是通过求解主光谱数据 X_m 和从光谱数据 X_s 间的转换矩阵 F 找到 2 种数据的线性关系,再将待迁移数据 X_i 乘以转换矩阵 F 完成数据迁移。但 PDS 使用限定宽度的窗口对光谱进行切割,使得每段光谱都能独立标准化。采用 PLS 回归求解转换矩阵 F 。

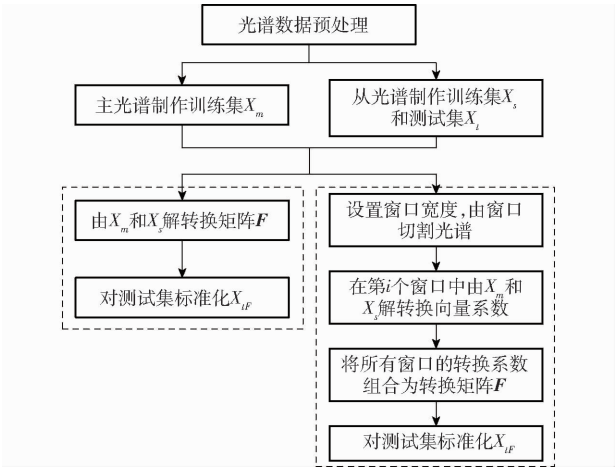


图 8 数据迁移算法流程图

Fig. 8 Flowchart of data migration algorithm

分别使用 2 种方法对测试集 30 个玉米样本的 NIR 传感器采集数据进行迁移。挑选 10 个含水率与训练集 X_m 接近的样本作为 X_s , 求解转换矩阵 F 。数据迁移效果对比如图 9 所示,为含水率 10.17% 玉米籽粒样本在波段 1 100 ~ 1 700 nm 内的吸收光谱。对比数据迁移效果可发现 PDS 数据偏移幅度更大,而 DS 数据迁移没有明显数据偏移。

2.4 含水率回归预测模型建立

对异常样本去除后的 70 个训练样本,采用梯度提升决策树(GBDT)训练回归模型并将 EVRI 数据集剩余的 10 个样本作为验证集。GBDT 有优秀的非线性关系预测能力和可解释性,常用于某种

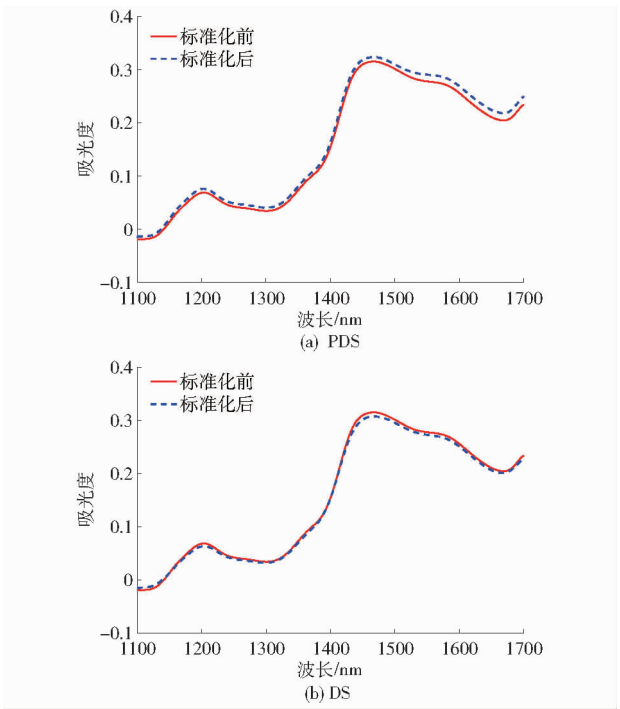


图 9 PDS 和 DS 对同一样本数据迁移效果对比

Fig. 9 Comparison of data migration effects of PDS and DS on the same sample

物质含量的预测问题^[24],GBDT 有 3 种常用损失函数,分别为平方差损失函数、绝对差损失函数、Huber 损失函数。平方差损失函数是 GBDT 最常用的损失函数之一,由于差值的平方关系因此对异常值敏感,从图 10a 可知,平方差损失函数能取得较好的预测结果,这是由于测试集经过去噪和模型迁移更加贴合训练集。绝对差损失函数相较于平方差损失函数其对大的误差惩罚较轻,因此对异常值相对不敏感,从图 10b 可知对于异常值较少的情况下其预测精度不如平方差损失函数。Huber 损失函数是平方差损失函数和绝对差损失函数的结合,通过参数 δ 控制,对于小误差,表现为均方误差(MSE),而对于大误差,则表现为平均绝对误差(MAE)。这使得 Huber 损失函数对异常值具有一定鲁棒性,同时保持了对较小误差的敏感性,从图 10c 可知,优化 δ 的 Huber 损失函数具有很高预测精度,此时 δ 为 0.8。以验证集决定系数 R_v^2 、相对分析误差(RPD_v)、预测均方根误差(RMSEP_v)、平均相对误差(MRE_v)对模型性能进行评价,结果如表 2 所示,可以发现 Huber 损失函数对于光谱类问题的回归预测有明显优势。

为判断模型对含水率预测是否具有可解释性,对特征重要性进行分析。图 11 为 GBDT 回归预测模型在 Huber 损失函数下贡献度最高的 10 个特征波长,可以发现在波段 1 480 ~ 1 676 nm 有较高贡献度,此波段属于水的第一倍频区^[25],是农作物含水

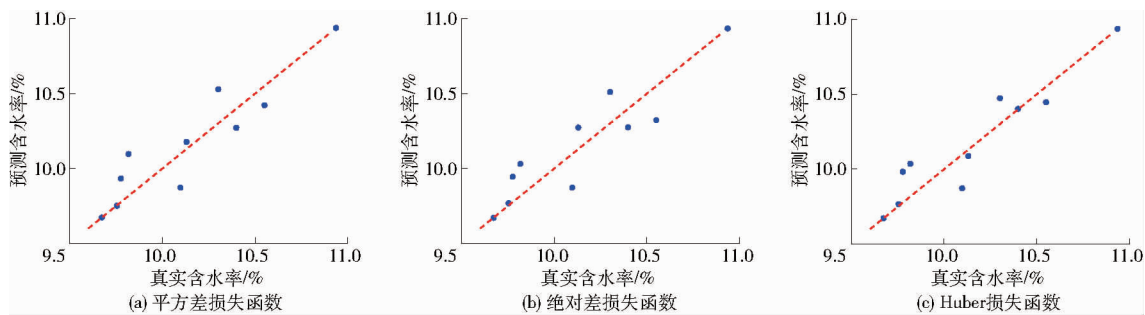


图 10 3 种不同损失函数预测结果

Fig. 10 Prediction results of three different loss functions

表 2 3 种不同损失函数预测结果评价指标

Tab.2 Evaluation indicators for prediction results of three different loss functions

损失函数	R^2_V	RPD_V	$RMSEP_V/\%$	$MRE_V/\%$
平方差	0.840 1	2.636 3	0.154 8	1.18
绝对差	0.828 7	2.547 0	0.160 2	1.31
Huber	0.877 7	3.014 8	0.135 4	0.98

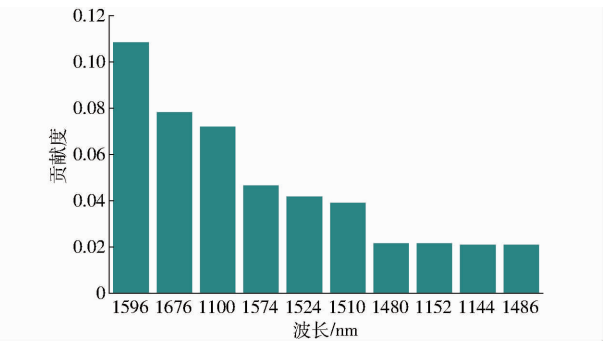


图 11 特征重要性分析

Fig. 11 Feature importance analysis

率分析最重要的区间。另外在波长 1 100 nm 对含水率也有较高贡献度,是因为 1 100 nm 在 C—H 键的第二信频区,能反映玉米籽粒油脂含量,而在自然生长的玉米籽粒中,含水率往往与成熟度、胚乳结构、油脂含量存在一定的统计相关性。

3 试验与结果分析

3.1 试验平台搭建

含水率检测系统如图 12 所示,整个系统由卤钨灯光源(HL-2000-HP-FHSA,美国)、光纤(QP400-2-VIS-NIR,美国)、暗盒、旋转台、NIR 传感器、Android 手机组成。其中光源光纤和接收光纤末端分别配有透镜固定在测量仓上层,用于聚焦照射光和漫反射光。暗盒内安装有旋转台,将样本放置在旋转台上即可实现样本的多点扫描。暗盒内壁贴有消光绒,以吸收反射到内壁上的杂散光。

3.2 样本制备

根据 GB/T 10362—2008《粮油检验 玉米水分

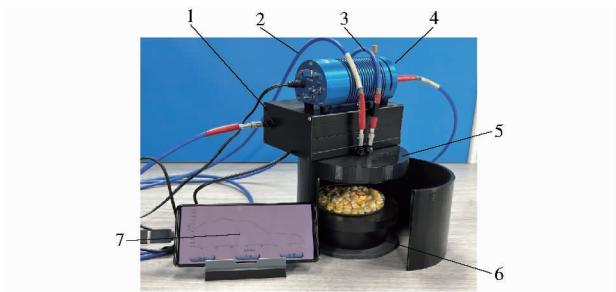


图 12 含水率检测系统

Fig. 12 Detection system of moisture content

1. NIR 传感器 2.接收光纤 3.光源光纤 4.光源 5.暗盒
6.旋转台 7.Android 手机

测定》方法测定样本的初始含水率,使用含水率测定仪(WTS-20A 型)测得郑单 958 玉米含水率为 12.60%。将此批次玉米分为 6 份,每份 1 kg。恒温干燥箱(LC-101-00B 型)设定为 90℃ 预热 30 min,将 3 份玉米分别平铺于金属托盘中并放入预热好的恒温干燥箱,分别干燥 70、80、90 min 后放在室温下冷却至常温。冷却至常温后将每个托盘中样本分别分装在 5 个密封罐中。然后按同样方法将另外 3 份玉米放入预热好的恒温干燥箱,分别干燥 100、110、120 min 后置于托盘在室温下冷却至常温。冷却至常温后将每个托盘中样本分别分装在 5 个密封罐中。共获得 30 分玉米样本。每个样本取 5.2~5.7 g 玉米,磨碎后取 5 g 玉米粉使用测定仪测量含水率。测得所有样本含水率为 9.73%~10.94%。

3.3 试验方法

测量仓内不放置任何样本,关闭光源,使用手机控制 NIR 传感器采集一组暗背景光谱数据。取 1 罐玉米样本倒入玻璃皿铺满,放在测量仓旋转台上,将反射白板放在玉米表面,采集一组白背景数据。拿下白板打开旋转台,每旋转 90°采集一次漫反射光谱,每个样本采集 4 次,并求得平均值即为当前样本光谱数据。根据比尔-琅勃定律计算吸光度。每个样本均计算吸光度并储存在手机中,通过应用程序中回归模型预测每个样本含水率。

3.4 结果分析

3.4.1 光学性能对比

目前市面上商业化的 NIR 传感器大多采用罗兰圆结构,其特殊的罗兰圆光栅可进一步缩小体积。但由于罗兰圆像面为弧形而探测器为线阵结构,无法保证所有波长均能在像面聚焦,直观的体现为分辨率较低。CT 结构传感器和罗兰圆结构传感器性能对比如表 3 所示,可以发现 CT 结构在分辨率上具有明显优势。

表 3 CT 结构和罗兰圆结构光谱传感器性能对比
Tab.3 Performance comparison between CT structure and Rowland circle structure spectrum sensor

光学结构	像素数	分辨率/nm	狭缝宽度/ μm
CT 结构	1 024	≤ 2	25
罗兰圆结构	128	≤ 16	60

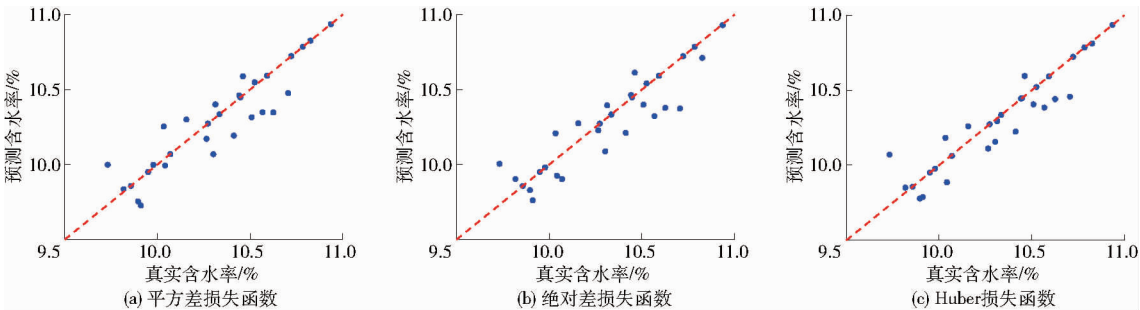


图 13 3 种 GBDT 模型测试集预测结果

Fig. 13 Prediction results of three GBDT models on test set

表 4 3 种 GBDT 模型预测结果评价指标
Tab.4 Evaluation indicators for prediction results of three GBDT models

损失函数	R^2_T	RPD_T	$RMSEP_T/\%$	$MRE_T/\%$
平方差	0.827 8	2.451 0	0.135 5	0.90
绝对差	0.821 8	2.409 7	0.137 8	0.95
Huber	0.861 7	2.734 6	0.121 4	0.78

定茶叶含水率,建立的 PLSR 模型 R^2 和 RMSEP 分别为 0.850 5 和 0.104 1%。LI 等^[27]在可见光至长波近红外波段(450 ~ 2 400 nm)对去皮的鲜核桃采样光谱,通过动力学模型预测含水率,其测试集 R^2 和 RMSE 分别为 0.821 和 1.877%。张静等^[16]分别使用便携式近红外(901 ~ 1 650 nm)和可见光(400 ~ 900 nm)光谱仪检测水稻含水率,其中近红外光谱采用结合 SPA 的 PLSR 模型, R^2 和 RMSEP 分别为 0.810 3 和 0.412%,可见光光谱采用结合 CARS 的 PCR 模型, R^2 和 RMSEP 分别为 0.802 1 和 0.388%。与现有 NIR 或 Vis - NIR 检测方法对比,本文快速检测平台具有良好检测精度。

常见的 2 种快速检测方法分别为微量烘干法和定容称量法,采用这 2 种方法分别测量 30 份玉米样

3.4.2 预测精度

将 30 份玉米样本光谱数据作为测试集,通过 SG 去噪和 PDS 数据迁移后得到平滑低噪声谱线。

由 GBDT 训练好的模型对测试集进行回归预测,结果如图 13 所示,并通过测试集 R^2_T 、 RPD_T 、 $RMSEP_T$ 、 MRE_T 对预测结果进行评价,结果如表 4 所示。当损失函数为 Huber 时,预测玉米含水率和真实含水率间误差最小,但 R^2_T 和 RPD_T 分别为 0.861 7 和 2.734 6,略低于前期验证集,这是由于去噪和模型迁移产生的数据失真对回归预测的影响。综合上述结果,在手机中内置 SG 去噪和 PDS 数据迁移,并采用 Huber 损失函数训练的 GBDT 模型,能够实现玉米含水率的有效预测。

3.4.3 其他检测方法对比

HUANG 等^[26]使用 Vis - NIR 光谱采集系统测

本含水率并对比本文含水率检测系统。表 5 为 3 种方法特性,综合对比可知本文检测平台具有精度高、检测速度快、对样本无损伤的特点。

表 5 3 种农作物含水率快速检测方法特性
Tab.5 Characteristics of rapid detection methods for water content of three crops

检测方法	绝对误差/%	样本检测速度/min	样本损伤
微量烘干法	0.1	40.0	有损
定容称量法	1.2	0.5	无损
NIR 光谱法	0.4	0.5	无损

4 结论

(1)通过 Zemax 对检测平台光学结构进行设计和优化,选定 CT 光学结构,其探测波段为 950 ~ 1 700 nm,整个检测范围内光学分辨率均可保持在 2 nm。该 NIR 传感器准直镜和聚焦镜均为离轴设计,表面镀银作为反射膜,通过 USB 与手机连接对 NIR 传感器进行控制,包括调节积分时间和连续测量等。

(2)确定了 NIR 传感器采样后的数据预处理流程。经过 SG 去噪,SNR 从 18.11 dB 提升至 35.10 dB,

PSNR 从 24.01 dB 提升至 39.94 dB,PDS 数据迁移算法模型预测准确率更高。

(3)以玉米含水率预测为例,GBDT 模型训练集为 70 组波段 1 100 ~ 1 700 nm 吸收光谱和玉米籽粒含水率,光谱间隔波长为 2 nm,并以 EVRI 提供数据中的 10 组样本作为验证集,通过验证集数据可知当损失函数为 Huber 时预测效果最好, R_v^2 、 RPD_v 、 $RMSEP_v$ 、 MRE_v 分别为 0.877 7、3.014 8、0.135 4%、0.98%。使用本文 NIR 传感器采集 30 组玉米数据作为测试集对模型进行测试,可知当损失函数为 Huber 时预测效果最好, R_T^2 、 RPD_T 、 $RMSEP_T$ 、 MRE_T 分别为 0.861 7、2.734 6、0.121 4%、0.78%。

(4)相较于其他快速检测方法,本文快速检测平台可以兼顾精度和速度,绝对误差小于等于 0.4%,单个样本检测速度为 0.5 min。

参 考 文 献

[1] 王小萌,吴文福,尹君,等. 玉米粮堆霉变发热过程中的温湿度场变化规律研究[J]. 农业工程学报, 2019, 35(3): 268 – 273.
WANG Xiaomeng, WU Wenfu, YIN Jun, et al. Research on temperature and humidity field change during corn bulk microbiological heating[J]. Transactions of the CSAE, 2019, 35(3): 268 – 273. (in Chinese)

[2] 李璐璐,薛军,谢瑞芝,等. 夏玉米籽粒含水率对机械粒收质量的影响[J]. 作物学报,2018,44(12):1747 – 1754.
LI Lulu, XUE Jun, XIE Ruizhi, et al. Effects of grain moisture content on mechanical grain harvesting quality of summer maize [J]. Acta Agronomica Sinica, 2018, 44(12): 1747 – 1754. (in Chinese)

[3] 江黎明,孙宗玖,汤丽斯,等. 收获期对伊犁绢蒿种子含水率、千粒重及萌发行为的影响[J]. 中国草地学报, 2019, 41(5): 7 – 16.
JIANG Liming, SUN Zongjiu, TANG Lisi, et al. Effects of harvest time on seed water content, 1000-grain weight and germination behavior of *Seriphidium transiliense* [J]. Chinese Journal of Grassland,2019,41(5):7 – 16. (in Chinese)

[4] 中华人民共和国国家质量监督检验检疫总局. 粮油检验 玉米水分测定:GB/T 10362—2008[S]. 北京:中国标准出版社, 2008.

[5] 河南中烟工业有限责任公司. 一种基于卤素水分仪的烟草含水率的检测方法: CN116086909A [P]. 2023 – 05 – 09.

[6] LU Y, HU Y, LI P. Consistency of electrical and physiological properties of tea leaves on indicating critical cold temperature [J]. Biosystems Engineering, 2017, 159: 89 – 96.

[7] CHIN-HASHIM N F, KHALED A Y, JAMALUDIN D, et al. Electrical impedance spectroscopy for moisture and oil content prediction in oil palm (*Elaeis guineensis* Jacq.) fruitlets[J]. Plants, 2022, 11(23): 3373.

[8] HAN C, WANG Y, SHI Z, et al. The design and experimentation of a corn moisture detection device based on double capacitors[J]. Sensors, 2024, 24(5): 1408.

[9] NATH K D, RAMANATHAN P. Non-destructive methods for the measurement of moisture contents-a review [J]. Sensor Review, 2017, 37(1): 71 – 77.

[10] VIRK S S, FULTON J P, FASINA O O, et al. Capacitance and near-infrared techniques for the real-time moisture measurement of broiler litter[J]. Biosystems Engineering, 2013, 116(4): 357 – 367.

[11] ZHANG Y, ZHAO J, GU Y, et al. A methodology study on the optimal detection of oil and moisture content in soybeans using LF – NMR and its 2D T1 – T2 nuclear magnetic technology[J]. Agronomy, 2023, 13(4): 1102.

[12] WEI Y, LI X, HE Y. Generalisation of tea moisture content models based on VNIR spectra subjected to fractional differential treatment[J]. Biosystems Engineering, 2021, 205: 174 – 186.

[13] MOON E J, KIM Y, XU Y, et al. Evaluation of salmon, tuna, and beef freshness using a portable spectrometer[J]. Sensors, 2020, 20(15), 4299.

[14] 刘文政,周雪健,平凤娇,等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. 农业机械学报, 2024, 55(2): 372 – 383.
LIU Wenzheng, ZHOU Xuejian, PING Fengjiao, et al. Detection of key indicators of ripening quality in table grapes based on visible – near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 372 – 383. (in Chinese)

[15] 彭彦昆,杨清华,王文秀. 基于近红外光谱的猪肉水分在线检测与分级[J]. 农业机械学报, 2018, 49(3): 347 – 353.
PENG Yankun, YANG Qinghua, WANG Wenxiu. On-line detection and classification of pork moisture based on near-infrared spectra[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(3): 347 – 353. (in Chinese)

[16] 张静,郭榛,王思花,等. 便携式近红外和可见光光谱仪检测水稻水分含量方法比较研究[J]. 光谱学与光谱分析, 2023, 43(7): 2059 – 2066.
ZHANG Jing, GUO Zhen, WANG Sihua, et al. Comparison of methods for water content in rice by portable near-infrared and visible light spectrometers[J]. Spectroscopy and Spectral Analysis, 2023, 43(7): 2059 – 2066. (in Chinese)

[17] 李扬裕,方勇华,李大成,等. 基于平板波导的小型红外光栅光谱仪光学设计[J]. 光谱学与光谱分析, 2015, 35(3): 841 – 845.
LI Yangyu, FANG Yonghua, LI Dacheng, et al. Optical design of miniature infrared gratings spectrometer based on planar

waveguide[J]. Spectroscopy and Spectral Analysis, 2015, 35(3): 841 – 845. (in Chinese)

[18] GATTINGER P, ZORIN I, RANKL C, et al. Spectral-coding-based compressive single-pixel NIR spectroscopy in the sub-millisecond regime[J]. Sensors, 2021, 21(16): 5563.

[19] JEON J, PARK J W, KIM G B, et al. Visible to near-infrared single pixel microspectrometer using electrothermal MEMS grating[J]. Optics Express, 2023, 31(9): 14583 – 14592.

[20] 高浩,王潇,尚林伟,等. 基于二向色及透射准直的小型近红外拉曼光谱仪[J]. 光谱学与光谱分析, 2018, 38(6): 1933 – 1937.

GAO Hao, WANG Xiao, SHANG Linwei, et al. Design and application of small NIR – Raman spectrometer based on dichroic and transmission collimating[J]. Spectroscopy and Spectral Analysis, 2018, 38(6): 1933 – 1937. (in Chinese)

[21] 陈潭轩,杨怀栋,陈科新,等. 宽光谱 Czerny – Turner 光谱仪中的彗差与分辨率[J]. 光谱学与光谱分析, 2010, 30(6): 1692 – 1696.

CHEN Tanxuan, YANG Huaidong, CHEN Kexin, et al. Coma and resolution in wide spectral region Czerny – Turner spectrometer[J]. Spectroscopy and Spectral Analysis, 2010, 30(6): 1692 – 1696. (in Chinese)

[22] ZHANG Y, HUANG Y, LI W, et al. Application of artificial neural network and DS algorithm to calibration transfer of rice protein powder[C]//2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), 2016: 822 – 825.

[23] WANG M, ZHENG K, YANG G, et al. A robust near-infrared calibration model for the determination of chlorophyll concentration in tree leaves with a calibration transfer method[J]. Analytical Letters, 2015, 48(11): 1707 – 1719.

[24] GARCÍA N P J, GARCÍA-GONZALO E, SÁNCHEZ L F, et al. Cyanotoxin level prediction in a reservoir using gradient boosted regression trees: a case study[J]. Environmental Science and Pollution Research, 2018, 25(23): 22658 – 22671.

[25] 樊梦丽,赵越,刘言,等. 近红外光谱水光谱组学[J]. 化学进展, 2015, 27(2/3): 242 – 250.

FAN Mengli, ZHAO Yue, LIU Yan, et al. Aquaphotomics of near infrared spectroscopy[J]. Progress in Chemistry, 2015, 27(2/3): 242 – 250. (in Chinese)

[26] HUANG Z, SANAEIFAR A, TIAN Y, et al. Improved generalization of spectral models associated with Vis – NIR spectroscopy for determining the moisture content of different tea leaves[J]. Food Engineering, 2021, 293: 110374.

[27] LI C, ZHANG S, SUN H, et al. Study on a two-dimensional correlation visible-near infrared spectroscopy kinetic model for the moisture content of fresh walnuts stored at room temperature[J]. Food Process Engineering, 2020, 43:e13551.

~~~~~

(上接第 298 页)

[21] LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized bert pretraining approach[J]. arXiv Preprint,arXiv:1907.11692,2019.

[22] DE MELLO H R, DE PAIVA A C, SILVA A C, et al. Application of natural language modeling techniques in natural gas segmentation in seismic reflection images[J]. Neural Computing and Applications, 2025, 37(4): 2383 – 2409.

[23] 杨璐,张恬,郑丽敏,等. 兽药致病命名实体 Att – Aux – BERT – BiLSTM – CRF 识别[J]. 农业机械学报,2022,53(3): 294 – 300.

YANG Lu, ZHANG Tian, ZHENG Limin, et al. Recognition of animal drug pathogenicity named entity based on Att – Aux – BERT – BiLSTM – CRF[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(3): 294 – 300. (in Chinese)

[24] 孙艳歌,吴飞,姚建峰,等. 多尺度自注意力特征融合的茶叶病害检测方法[J]. 农业机械学报,2023,54(12):308 – 315.

SUN Yange, WU Fei, YAO Jianfeng, et al. Tea disease detection method with multi-scale self-attention feature fusion[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(12):308 – 315. (in Chinese)

[25] 廖天正,林晓兰,陈永辉,等. 基于 BiLSTM – CRF 的命名实体识别在临床电子病历的研究与应用[J]. 电脑知识与技术, 2024,20(28):17 – 19.

[26] 程名,于红,冯艳红,等. 融合注意力机制和 BiLSTM + CRF 的渔业标准命名实体识别[J]. 大连海洋大学学报,2020, 35(2):296 – 301.

[27] YU B, WEI J. IDCNN – CRF-based domain named entity recognition method[C]//2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, 2020:542 – 546.